



UNIVERSITÉ
DE MONTPELLIER



Master Mention IGT Ingénierie et gestion territoriale et Mention
Informatique - Spécialité Geomatique
Université Montpellier II et Montpellier III

VERS UNE MÉTHODE OPÉRATIONNELLE DE CARTOGRAPHIE DE L'OCCUPATION DU SOL ET DES CULTURES À LA RÉUNION PAR ANALYSE ORIENTÉE OBJET D'IMAGES À TRÈS HAUTE RÉSOLUTION SPATIALE ET FOUILLE DE DONNÉES

Marie PISSON-GOVART

Soutenance le 21 Septembre 2015

Tuteur universitaire
Laurent DEMAGISTRI
Chercheur UMR ESPACE-DEV

Maitre de stage
Pierre TODOROFF
Chercheur CIRAD

Pierre TODOROFF

Fonction du tuteur : chercheur

Coordonnées du tuteur : CIRAD, 11 rue Paradis, 97410 Saint-Pierre - Réunion

N° Tel, fax et e-mail : Tél : 0262 49 92 61 ; Fax : 0262 49 92 95 ; pierre.todoroff@cirad.fr

Conditions du stage

Lieu et descriptif rapide de la structure d'accueil :

Centre de Coopération internationale en Recherche Agronomique pour le Développement (CIRAD).

Le CIRAD est un centre de recherche français qui répond, avec les pays du Sud, aux enjeux internationaux de l'agriculture et du développement.

Le/la stagiaire sera accueilli(e) dans l'Unité de Recherche « Agroécologie et intensification durable des cultures annuelles » (AIDA) qui mène des recherches sur l'intensification et la durabilité de la production des cultures annuelles en milieu tropical contraint. Ses recherches visent la pleine valorisation des ressources disponibles, en mobilisant les processus écologiques qui régissent leur dynamique au sein des agrosystèmes.

Plus spécifiquement le/la stagiaire effectuera son stage au sein de l'équipe ARTISTS (Télédétection, systèmes d'information, techniques de simulations et analyses spatiales), spécialisée dans l'analyse de l'information spatiale, basée à La Réunion. Cette équipe est constituée d'une dizaine de chercheurs, doctorants, techniciens spécialistes de la télédétection, des systèmes d'information, de la fouille de données et des modèles de cultures. Il/elle bénéficiera de leur appui et des infrastructures de l'équipe.

Equipements géomatiques :

Postes informatiques sous Windows, serveur de calcul, ressources réseau (sauvegarde, logiciels, ...), GPS différentiels, logiciels SIG (Trimble Pathfinder Office, ArcGIS) et de télédétection (Erdas Imagine, eCognition), de calcul numérique (Matlab, R), traceur A0.

Il/elle aura accès aux images de l'antenne de réception satellitaire SEAS-OI, aux archives d'images satellites de l'équipe (plusieurs centaines d'images de divers satellites en optique et radar, dont la couverture Pléiades de La Réunion).

Dates et durée : du 23 mars au 31 août 2015

Rémunération : 508,22€/mois

Nom de l'enseignant/chercheur chargé du suivi du stage (ou tuteur pédagogique) :

Fonction (PR, MCF, chercheur...) : Laurent DEMAGISTRI. Fonction: IE

Coordonnées UMR_ESPACE-DEV

Equipe Observation Spatiale de l'Environnement (OSE)
Maison de la télédétection, 500 rue Jean François Breton,
34 093 Montpellier Cedex 5

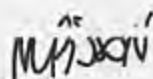
N° Tel, fax et e-mail : +33(0)4.67.54.87.01 ; laurent.demagistri@ird.fr

Signatures

Le stagiaire

L'enseignant/chercheur

Le tuteur en entreprise




Resp. Master Géomatique
 JP BORG



REMERCIEMENTS

Je remercie tout particulièrement mon tuteur Pierre TODOROFF pour son accueil et le temps qu'il m'a consacré lors de mon stage que j'ai eu la chance et le plaisir de passer au sein du CIRAD. Merci également pour la confiance qu'il m'a témoigné, sa patience et sa gentillesse.

Je tiens aussi à remercier l'ensemble de l'équipe ARTIST pour son accueil et sa gentillesse envers moi, ils m'ont permis de passer un stage dans la joie et la bonne humeur !

Je souhaite également remercier les professeurs des universités Montpellier II et Montpellier III pour les enseignements et le suivis pédagogique de ces deux années de Master.

INTRODUCTION	1
I. CONTEXTE DE L'ÉTUDE	2
SITUATION GÉOGRAPHIQUE DE L'ÎLE DE LA RÉUNION	2
Le relief	2
Démographie	3
Une agriculture ambitieuse	3
LE CIRAD À LA RÉUNION	4
Dans le monde	5
À Saint-Pierre, La Réunion	5
L'équipe ARTISTS 5	
LA TÉLÉDÉTECTION POUR LA CARTOGRAPHIE DE L'OCCUPATION ET DE L'UTILISATION DU SOL	7
L'analyse orientée pixel	7
L'analyse orientée objet	8
Première étape : la segmentation	9
Deuxième étape : la classification des segments	10
Random Forest	12
Support Vector Machine SVM	13
Définition de la typologie des classes	14
II. MATÉRIEL ET MÉTHODE	15
IMAGES SATELLITES ET DONNÉES GÉOGRAPHIQUES	15
Données Raster	15

Données Vecteur	16
PRÉPARATION DES DONNÉES	17
Images Pléiades	17
MNT Litto 3D et extraction des zones de fortes pentes	18
Données vecteur	19
BASE DE DONNÉES	20
Variables descriptives	20
Jeu de données	20
ADAPTATION DU MODÈLE DE 2014	21
Segmentation des images	22
Classification des objets	24
Extraction des classes générales d'occupation du sol par règles expertes	24
Extraction des classes de végétation par algorithme de fouille de données	25
<i>Classification des classes végétation naturelle et Surfaces agricoles en culture</i>	25
<i>Classification des classes canne à sucre, verger, prairie, maraichage et diversification</i>	26
TEST DE REPRODUCTIBILITÉ SUR D'AUTRES RÉGIONS	28
III. ANALYSE DES RÉSULTATS	29
ADAPTATION DU MODÈLE SUR LA ZONE DE SAINT-PIERRE À D'AUTRES DATES	29
Segmentation des images	29
Classification des objets	29
Classification des classes générales d'occupation du sol	29

Extraction des classes d'utilisation du sol par algorithme de fouille de données	31
<i>Classification des classes Végétation naturelle et Surfaces agricoles en culture</i>	31
<i>Classification des cultures</i>	32
TEST DE REPRODUCTIBILITÉ SUR D'AUTRES RÉGIONS	33
Segmentation des image	33
Classification des objets	33
Classification des classes générales d'occupation du sol	33
Extraction des classes d'utilisation du sol par algorithme de fouille de données	34
<i>Classification des classes Végétation naturelle et Surfaces agricoles en culture</i>	34
<i>Classification des cultures</i>	35
IV. DISCUSSION	36
LIMITES ET PERSPECTIVES	36
ÉVOLUTIONS POSSIBLES	36
V. CONCLUSION	38

LISTE DES FIGURES

Figure 1 : Situation géographique de l'île de L Réunion	2
Figure 2 : Implantation des directions régionales du CIRAD dans le monde	5
Figure 3 : Illustration du type de fonction d'appartenance Gaussien	11
Figure 4 : Illustration du fonctionnement de l'algorithme Random Forest	12
Figure 5 : Exemple d'application des SVM	13
Figure 6 : Aperçu des images Pléiades couvrant l'ensemble du territoire	16
Figure 7 : Modèle Builder permettant l'extraction des pentes sur les zones étudiées	19
Figure 8 : Aperçu des images satellites couvrant la zone de Saint-Pierre	21
Figure 9 : Synthèse des traitements successifs de segmentation / classification	27
Figure 10 : Images Pléiades traitées sur le reste du territoire	28
Figure 11 : Aperçu des objets issus des segmentations de niveau 1 et 2	29
Figure 12 : Aperçu des classification du Tissu urbain et des Surfaces bâties après ajustement des seuils	30
Figure 13 : Comparaison des coefficient de Kappa obtenus pour les classifications avec l'algorithme des Random Forest pour l'ensemble des images couvrant la zone de Saint Pierre	31
Figure 14 : Comparaison des coefficients de Kappa obtenus pour les modèles de classification élaborés avec l'algorithme des SVM	33

LISTE DES TABLEAUX

Tableau 1 : Typologie des classes	14
Tableau 2 : Récapitulatif des paramètres de segmentation utilisés	22
Tableau 3 : Seuils retenus pour extraire les classes d'occupation du sol par règles expertes	24
Tableau 4 : Test de transfert des modèles établis à partir de chaque image	32
Tableau 5 : seuils de classification utilisés pour l'extraction des classes générales d'occupation du sol	34
Tableau 6 : Matrice de confusion et précision globale de la classification par l'algorithme des Random Forest sur la zone Nord de l'image du 02/09/2013	34
Tableau 7 : Matrice de confusion et précision globale de la classification par l'algorithme des Random Forest sur la zone Nord de l'image du 02/09/2013	35

INTRODUCTION

La connaissance des ressources naturelles et agricoles terrestres, aussi bien agraires que forestières est une préoccupation primordiale pour l'homme depuis toujours.

Ainsi, l'Île de la Réunion, en fort développement démographique et économique, fait face à des pressions foncières considérables dues à la réduction des espaces disponibles. L'aménagement difficile de certaines régions de l'île à cause du relief très accidenté du territoire contraint davantage les scénarii de développement possibles. Pour orienter au mieux les stratégies de développement territoriale, les acteurs du territoire ont besoin de connaître avec le plus de précision possible la nature de l'occupation du sol.

Une multitude d'études montre le fort potentiel de l'analyse d'images satellites pour la cartographie de l'occupation du sol. Les avancées technologiques dans le domaine de la télédétection et des ressources informatiques fournissent aujourd'hui des outils et des données suffisamment précises pour étudier l'occupation des sols à des échelles toujours plus fines. En effet, l'utilisation des images satellites à très haute résolution spatiale (inférieure à 1m) couplée à de nouvelles techniques d'analyse telle que l'analyse orientée objet permet d'obtenir une cartographie de l'occupation du sol avec des niveaux de détails jusqu'ici inégalés.

L'étude présentée ci-après consiste à affiner et valider une méthode de classification de l'occupation et de l'utilisation du sol par analyse orientée objet d'images à très haute résolution spatiale développée en 2014 sur une zone test et la traduire en une méthode de cartographie opérationnelle valable sur l'ensemble du territoire. Le rapport présente en premier lieu, le contexte dans lequel s'inscrit cette étude, le perfectionnement de la méthode existante sur la zone d'étude de Saint Pierre et se concentre ensuite sur la généralisation de cette méthodologie pour son utilisation à partir d'autres dates d'acquisition d'images et sur d'autres régions de l'île. Finalement, le bilan relatif à ce stage propose des perspectives d'évolution du modèle et souligne les limites et les apports personnels de ce travail.

I. CONTEXTE DE L'ÉTUDE

SITUATION GÉOGRAPHIQUE DE L'ÎLE DE LA RÉUNION

La Réunion, d'une superficie de 2512km², est une île de l'archipel des Mascareignes située dans l'océan Indien à environ 700km à l'est de Madagascar, plus exactement localisée à 21° de latitude Sud et 55° de longitude Est en pleine zone tropicale. Ancienne colonie française, La Réunion est devenue un département français en 1946 entraînant un rapide essor socio-économique, et plus récemment, dans les années 90, une région ultra-périphérique de l'Union Européenne lui permettant de bénéficier d'aides européennes.

Depuis 2010, les « Pitons, cirques et remparts de la Réunion » sont classés au Patrimoine Mondial de l'UNESCO assurant à l'île une reconnaissance internationale et une renommée offrant de belles perspectives de développement touristique. Ce parc national couvre une superficie de 1024km², réduisant de moitié les espaces aménageables de l'île (environ 40% du territoire insulaire).



Figure 1 : Situation géographique de l'île de La Réunion

Le relief

L'île de la Réunion est née suite à l'activité d'un point chaud de la plaque Africaine il y a quelques trois millions d'années avec l'émergence d'un massif montagneux appelé

le Piton des neiges qui culmine aujourd'hui à 3 070m (plus haut sommet de l'océan indien). Ce piton constitue la partie ouest de l'île, la partie est étant formée d'un second volcan, vieux de seulement cinq cent mille ans : le piton de la fournaise, toujours en activité.

L'espace insulaire est donc marqué par un relief très accidenté caractérisé par des pentes importantes de 6° à 8° en moyenne et pouvant s'élever jusqu'à plus de 30° (sur environ 20% du territoire). Ce relief contraint très fortement l'aménagement de l'île, les espaces disponibles devenant de plus en plus rares. Zones urbaines, activités industrielles, portuaires et touristiques se concentrent le long du littoral entre 0 et 500 mètres d'altitude (les « Bas ») et occupent principalement les 40% du territoire aménageables. Les «Hauts» (plus de 500m d'altitude) sont constitués de zones à dominante rurale consacrées à l'élevage et à la conservation du patrimoine.

Démographie

La Réunion connaît, depuis sa départementalisation, une accélération de ses dynamiques spatiales liée à un essor socio-économique rapide (Chapelain, 1993). La croissance de ses espaces urbains figure au premier rang de ses préoccupations d'aménagement. En effet, la Réunion fait face à un taux annuel de croissance de la population spectaculaire : +1,5% par an, soit trois fois plus qu'en métropole. Selon les projections de l'INSEE, l'île comptera 1 million d'habitants en 2030. Parallèlement à la croissance des espaces urbains (principalement sur le littoral et les mi-pentes), on observe une réduction notable des espaces agricoles : 800 ha de moins par an depuis 2007 (INSEE).

Une agriculture ambitieuse

Bien que l'île soit sujette à des tempêtes cycloniques régulières pendant l'été une grande variété de produits est cultivée sur le territoire, les micro-climats localisés permettant une grande diversité de cultures : canne à sucre, horticulture, fruit, légume, viticulture, etc....

L'agriculture occupe une place importante dans l'économie de l'île, les exportations représentent au total 286,9 millions d'euros dont 69,7 millions proviennent du secteur sucrier. Au total, 43 313 hectares sont consacrés au domaine agricole dont 57% dédiés à la filière canne à sucre. En effet, la canne est la première production agricole de l'île. Cette culture présente de nombreux avantages : résistance aux aléas climatiques, production de sucre et de rhum, et, production d'énergie à partir

de la biomasse. Ce nouveau débouché vient alimenter les tensions en exacerbant la concurrence entre les filières agricoles qui doivent maintenir un niveau de production élevé sur des espaces de plus en plus réduits.

De plus, le projet D.E.F.I. : Développement de l'Élevage et des Filières des Interprofessionnels porté par les associations ARIBEV (Association Réunionnaise Interprofessionnelle pour le Bétail Et la Viande) et ARIV (Association Réunionnaise Interprofessionnelle de la Volaille) avec le soutien de l'État et de l'Union Européenne impacte fortement les espaces agricoles. En effet, ce projet débuté en janvier 2011 à comme objectifs de réduire les coûts des produits locaux, augmenter les emplois (+1600 emplois locaux directs ou indirects), participer à l'installation de nouveaux éleveurs (+140), augmenter la production locale et réduire au maximum la dépendance alimentaire de l'île. Soutenu par les différents acteurs et par la population, le projet participe également à la réduction des espaces disponibles et installe ainsi une forte compétition entre les filières agricoles et animales.

LE CIRAD À LA RÉUNION

Dans le monde

Le CIRAD, Centre de coopération Internationale de Recherche Agronomique pour le Développement, est un centre de recherche français placé sous la tutelle conjointe du Ministère de la Recherche et du Ministère des Affaires Etrangères qui oeuvre pour le développement durable des régions tropicales et méditerranéenne. Il a pour missions de produire et de transmettre des connaissances afin d'accompagner le développement agricole des pays en développement et de participer aux grands débats sur les enjeux agricoles et alimentaires mondiaux.

Le CIRAD se compose de 12 directions régionales, en plus des directions générales de Languedoc Roussillon et d'Ile-de-France, réparties entre l'outre-mer français et différentes régions du monde (cf. carte ci-dessus).



Figure 2 : Implantation des directions régionales du CIRAD dans le monde

À Saint-Pierre, La Réunion

À la Réunion le Cirad travaille en partenariat avec l'Etat et les collectivités locales, il participe aux réflexions de la recherche internationale sur la sécurité alimentaire, le développement durable et l'adaptation de l'agriculture au changement climatique.

Le CIRAD est implanté sur quatre stations de recherche : une à Saint-Denis au nord, deux à Saint-Pierre (Ligne Paradis et Bassin plat) au sud et une à Saint-Leu à l'ouest de l'île. Ces stations interviennent principalement dans trois domaines :

- Protection des plantes
- Risque environnemental, agriculture et gestion intégrée des ressources
- Qualité des productions agricoles et des produits alimentaires

L'équipe ARTISTS

L'équipe ARTISTS, Artists of Remote sensing Tools, Information systems, Simulation Techniques and Spatial analysis, travaille sur la production d'outils d'aide à la décision par analyse d'information spatiale. Ces outils s'appuient principalement sur des modèles agronomiques, des systèmes d'information géographique et des données issues de la télédétection pour élaborer des outils d'aide à la décision destinés aux acteurs des filières agricoles.

L'étude présentée ici a été réalisée au sein de cette équipe. Depuis 2014, l'équipe travaille sur la problématique de la cartographie de l'occupation et de l'utilisation du

sol de l'île : deux projets de recherches sont développés : un premier utilisant des données radar et l'autre, utilisant des images satellites à très haute résolution spatiale (THRS). L'objectif de ces études est de mettre en place une méthode opérationnelle de cartographie de l'occupation et de l'utilisation du sol à partir de données issues de la télédétection.

Une première étude a été réalisée en 2014 étudiant le potentiel des images THRS Pléiades couplées à l'analyse orientée objet pour cartographier l'occupation et l'utilisation du sol sur une zone test. La méthodologie a été mise en place à partir d'une image Pléiades (0,5m de résolution) de juillet 2012 qui couvre la zone de Saint-Pierre, dans le sud de l'île. Elle utilise l'analyse orientée objet pour extraire 8 classes d'occupation du sol : 4 classes correspondent aux surfaces artificialisées et naturelles génériques et 4 classes correspondent aux surfaces agricoles en culture. Trois méthodes de classification différentes ont été utilisées pour discriminer ces classes :

- Une classification par règles expertes permet de classer le réseau routier, les surfaces urbaines, le sol nu et la végétation.
- Une classification supervisée utilisant l'algorithme des maximums de vraisemblance permet de différencier les classes végétation naturelle et surfaces agricoles.
- Une classification basée sur une analyse en composante principale et sur la logique floue permet de discriminer les classes culturelles à savoir la canne à sucre, les vergers, les prairies et le maraichage et diversification.

Cette méthode fournit des résultats encourageants : la végétation naturelle et le domaine agricole sont séparés avec une précision globale de 87% et les différentes classes culturelles avec une précision globale de 81%.

Dans ce contexte et dans la continuité de l'étude réalisée en 2014 qui montre le fort potentiel de l'analyse orientée objet d'images satellites à très haute résolution, l'étude présentée ici consiste à :

- mettre en place une méthodologie permettant de cartographier l'occupation et l'utilisation du sol par analyse orientée objet
- perfectionner la méthode expérimentale de 2014 dans le sud de l'île pour la rendre opérationnelle à d'autres dates

- généraliser et valider la méthode mise en place sur la zone de Saint-Pierre à d'autres régions et à d'autres dates
- améliorer la performance et adapter la méthode en vue d'une mise en oeuvre opérationnelle.

LA TÉLÉDÉTECTION POUR LA CARTOGRAPHIE DE L'OCCUPATION ET DE L'UTILISATION DU SOL

La télédétection se définit comme la capacité d'observation de la Terre à distance. Elle se base sur les radiations électromagnétiques émises, réfléchies, ou diffusées par la surface terrestre. Les techniques de télédétection se sont développées, ces dernières années, de manière spectaculaire et trouvent aujourd'hui une application à presque tous les domaines des sciences de la Terre.

Ces avancées technologiques récentes et le développement des ressources informatiques de traitement d'images satellite offrent aujourd'hui aux utilisateurs une gamme de produits très étendue permettant des études variées à grandes et/ou à petites échelles. Les nouvelles données à très haute résolution spatiale (taille du pixel inférieure à 1m) permettent d'établir une cartographie de l'occupation du sol toujours plus précise basée sur des informations de texture, de forme ou de géométrie mises en évidence grâce au très haut niveau de détails fournis par ces nouveaux satellites. Deux modes d'analyse existent : celle dite orientée pixel et celle dite orientée objet.

L'analyse orientée pixel

Il existe une multitude d'études basées sur les techniques de télédétection permettant de discrétiser différentes classes d'occupation du sol. La majorité de ces classifications d'images satellites se basent sur l'information de chaque pixel de l'image comme unité de base de l'information. Cette approche de classification conventionnelle se base sur l'assignation d'une classe thématique définie aux pixels étudiés individuellement et est donc dite orientée pixel. L'information spectrale de chaque pixel est ainsi utilisée comme base numérique pour l'assignation d'une classe (Alban, 2005). L'analyse d'image orientée pixel peut être :

- supervisée : les pixels sont associés à une classe thématique d'après une base de connaissance - données terrain - préalablement établie par le télédécteur, ou

- non-supervisée : les pixels sont classés automatiquement d'après leurs propriétés spectrales et suivant le nombre de classes défini par l'utilisateur.

Divers algorithmes sont disponibles pour ces deux types de classification, les plus utilisés (plus performants) étant le cluster ISODATA (Yale University, *Landcover Classification Project*) pour les classifications non-supervisées et l'algorithme des maximums de vraisemblance pour la classification supervisée.

Cette méthode d'analyse a prouvé son efficacité dans de nombreuses études. Cependant, elle se révèle peu performante pour la cartographie de paysages agricoles morcelés, très hétérogènes cultivés par des exploitations familiales comme c'est le cas à la Réunion. De plus, elle est limitée par son incapacité à prendre en compte le contexte et par l'apparition d'un effet « poivre et sel » sur les images à haute et très haute résolution spatiale (Blaschke, 2009) dû à l'hétérogénéité des pixels. Cette méthode est plus adaptée pour caractériser l'occupation du sol de zones homogènes.

De nouveaux objets jusque là non détectables sont maintenant percevables sur les images THRS et perturbent les méthodes de détection classique adaptées à l'analyse des images à résolution décimétrique (Puissant, 2006). En effet, selon Baatz et Schäpe : «*L'information sémantique précieuse pour interpréter une image, n'est pas reflétée dans les pixels individuels, mais dans un groupe de pixels qui représentent des objets significatifs et dans leurs relations mutuelles*» (Baatz et Schäpe 2000).

L'analyse orientée objet

Pour palier à ces problèmes inhérents à la très haute résolution spatiale, une nouvelle approche d'analyse d'image a été développée. Cette dernière se base sur l'identification préalable d'objets, définis comme des groupes de pixels contigus avec des caractéristiques similaires.

Ces objets sont créés selon un processus de segmentation qui rassemble des pixels en groupes (objets) homogènes, de manière à ce que la variabilité entre les pixels qui compose l'objet soit inférieure à la variabilité entre l'objet et son entourage. Ils servent ensuite d'unité élémentaire d'information pour la classification ultérieure. Les objets ainsi créés ne sont plus seulement caractérisés par leurs caractéristiques spectrales mais aussi par leurs caractéristiques de forme, de texture, etc. Dans ce cas, l'analyse

de l'image n'est plus basée sur les pixels mais sur les objets. De nombreux travaux démontrent l'intérêt de cette méthode en comparaison à la classification purement spectrale, centrée sur le pixel.

La principale force de l'analyse orientée objet est de pouvoir combiner plusieurs indices : réflectance, texture, géométrie, voisinage, information thématique, etc. pour isoler au mieux les différents types d'occupation du sol.

L'analyse orientée objet se compose de deux grandes étapes successives : la segmentation de l'image en objets homogènes et la classification des objets obtenus.

Première étape : la segmentation

La segmentation subdivise l'image en régions. Ce processus peut fournir un grand nombre de solutions différentes, la meilleure étant celle qui sépare le mieux les entités observées en fonction des critères recherchés.

Le processus de segmentation réduit la variance intra-classe et peut augmenter la précision de la classification si les paramètres sont bien choisis (Blaschke 2003, Addink et al. 2007). Néanmoins à ces résolutions, le nombre d'objets à identifier et leur hétérogénéité s'accroît ce qui complique et prolonge les procédures d'extraction des objets (Sparfel et al. 2008 ; Baatz et al. 2008).

Plusieurs algorithmes de segmentation sont disponibles dans le logiciel eCognition sur lequel nous travaillons :

- «Chessboard» qui segmente l'image suivant un maillage régulier de taille définie par l'utilisateur,
- «QuadTree» qui segmente l'image en regroupant les pixels homogènes dans des carrés de taille différente. Ici l'utilisateur définit un paramètre d'échelle,
- et enfin, «Multirésolution» qui segmente l'image en agglomérant des pixels homogènes jusqu'à atteindre un certain niveau d'hétérogénéité défini par l'utilisateur. Nous utiliserons cet algorithme dans la suite de l'étude.

La segmentation multi-résolution permet, pour différentes échelles de segmentation, de représenter l'information de l'image en différentes résolutions spatiales simultanément grâce à une organisation hiérarchique permettant une analyse de

l'image multi-échelle. Il s'agit d'une technique ascendante de croissance qui débute par un ou plusieurs objets constitués d'un seul pixel de l'image, les plus petits objets sont ensuite fusionnés de façon itérative jusqu'à atteindre le seuil d'homogénéité fixé par l'utilisateur : le processus s'arrête lorsque l'homogénéité des objets dépasse l'hétérogénéité maximale acceptée.

Pour obtenir des objets adaptés à l'étude, l'utilisateur fixe les critères d'homogénéité, les bandes spectrales utiles et le seuil d'hétérogénéité maximale. Il est possible d'attribuer un poids aux différentes bandes de l'image : plus le poids est élevé plus son information est utilisée pendant le processus. L'opérateur peut également importer des couches thématiques (données vectorielles - shapefile) et contraindre la segmentation à suivre les limites de celles-ci, en revanche, la pondération de ces couches n'est pas possible. L'opérateur définit ainsi le paramètre d'échelle, qui détermine l'hétérogénéité maximale permise, il influence directement la taille des objets issus de la segmentation. Les critères d'homogénéité des objets sont définis par les paramètres :

- «shape», directement lié à la couleur et la forme, il correspond à l'homogénéité spectrale des objets : si le critère est fixé à 0, les objets seront optimisés uniquement par leurs homogénéité spectrale) et
- « compactness » influence les contours des objets, un critère à 0 implique que les objets auront un contour très découpés.

Deuxième étape : la classification des segments

Avant de commencer la classification des objets, il est nécessaire d'établir une base de données renseignant les caractéristiques spectrales, texturales, spatiales et contextuelles de chaque segment. Ces attributs sont ensuite utilisés dans la construction des règles de classification.

Différentes alternatives existent pour établir une classification des objets issus des segmentations, nous utilisons des méthodes de classifications supervisées qui peuvent être de deux types :

- l'approche supervisée qui nécessite des échantillons d'apprentissage de référence et un ensemble d'attributs discriminants. Cette approche est propice dans le cas où l'assignation d'un objet à une classe particulière repose sur un grand nombre d'attributs.

- l'approche supervisée par construction de règles expertes qui applique des règles fixées par l'opérateur suivant ses observations et ses connaissances. Ces règles sont combinées grâce à des opérateurs logiques. Cette approche est plus adaptée aux cas où les objets peuvent être caractérisés par un nombre limité d'attributs.

Ces règles peuvent être basées sur une logique booléenne où les objets appartiennent ou non à la classe thématique définie (d'après les seuils fixés) ou alors sur une logique floue où, à chaque objet, est associé des degrés d'appartenance aux différentes classes thématiques, la classe ayant le plus haut degrés d'appartenance est celle retenue. Le modèle de fonction d'appartenance le plus communément utilisé est celui de type Gaussien (cf figure ci-dessous).

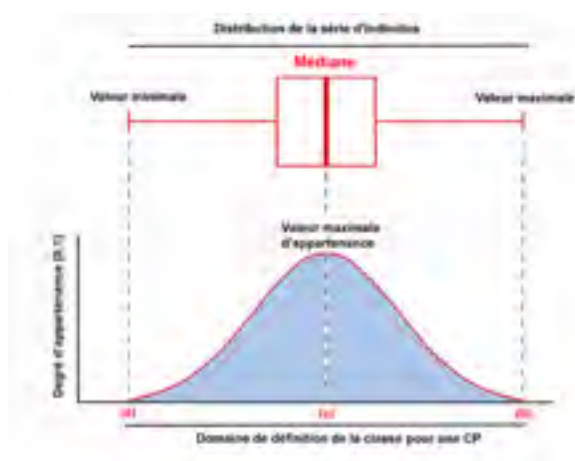


Figure 3 : Illustration du type de fonction d'appartenance Gaussien

Dans de récentes études, la supériorité des algorithmes d'apprentissage automatique comparé aux algorithmes classiques tels que celui des Maximum de vraisemblance (Maximum Likelihood) est mise en avant (Dixon et Candade, 2008; Yang et al., 2011; Nitze, Schultness, Asche 2012). Les précisions de classification d'algorithmes basés sur les arbres de décision tels que les Random Forest (ou forêts aléatoires), les réseaux de neurones et les Séparateurs à vaste marge (SVM) sont similaires. Dans l'étude menée par Dixon et Candade (2008), les résultats de classification sont semblables pour les SVM et les réseaux de neurones, en revanche, l'algorithme des Maximum de vraisemblance obtient de mauvais résultats. De même, dans l'étude menée par Nitze, Schultness et Asche en 2012 qui compare les performances des algorithmes SVM, réseau de neurones, Random Forest et Maximum Likelihood, les algorithmes SVM et Random Forest ressortent comme les

plus performants avec en moyenne, une précision globale de classification de 5% plus élevée que celle obtenue par les Maximum de vraisemblance. Dans la suite de l'étude, nous testerons les algorithmes des Random Forest et des séparateurs à vaste marge pour les classifications des classes de végétation.

Random Forest

Random Forest ou forêts aléatoires est une technique de classification basée sur les arbres de décision. Elle a été développée par Léo Breiman et Adèle Cutler en 1984. Il s'agit d'une technique d'apprentissage automatique basée sur la construction d'un arbre de décision discriminant les échantillons en différents groupes leur attribuant une classe. Ces regroupements sont construits à partir de règles s'appuyant sur des variables descriptives pouvant être quantitatives ou qualitatives.

Cette technique a été réadaptée (Breiman, 1996) pour améliorer les performances prédictives en construisant non plus un mais plusieurs dizaines d'arbres dont les réponses sont combinées pour obtenir les prédictions d'appartenance à une classe

(cf. figure 4).

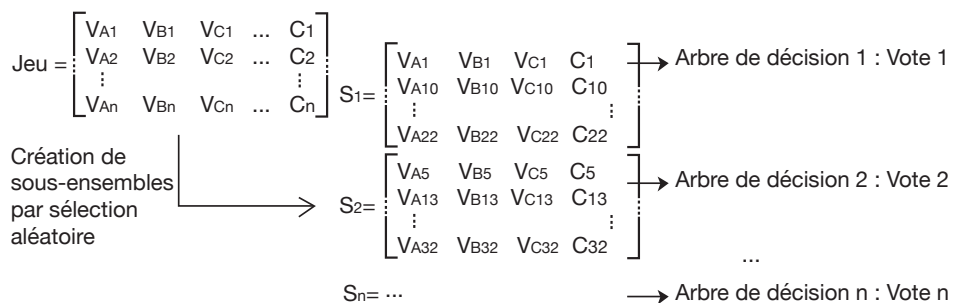


Figure 4 : Illustration du fonctionnement de l'algorithme Random Forest

En effet, l'algorithme des forêts aléatoires se base sur le constat que la combinaison de plusieurs modèles d'apprentissage accroît la précision des prédictions de classification. Ainsi, l'algorithme des forêts aléatoires fonctionne à partir d'un ensemble d'arbres de décision non corrélés : l'algorithme génère des sous-ensembles à partir d'une sélection aléatoire des entités étudiées. Pour chaque sous ensemble, un arbre de décision est généré qui émet un vote. A la fin, on récupère toutes les prédictions établies par tous les arbres de décision et on compte le nombre de vote pour chaque classe. La classe ayant le plus grand nombre de vote est celle retenue.

Cet algorithme présente de nombreux avantages :

- il est très efficace sur de grandes bases de données : il peut gérer des centaines de variables sans en exclure aucune et prend en compte les données hétérogènes,
- il détermine également les variables d'importance c'est à dire celle qui sont considérées comme les plus discriminantes pour les classes étudiées.
- il est implémenté dans les logiciels R et Weka.

Les paramètres à fixer pour exécuter l'algorithme sont le nombre d'arbres de décision souhaité et le nombre de variables à prendre en compte dans chacun d'entre eux.

Support Vector Machine SVM

Les SVM ou séparateurs à vaste marge, sont des techniques d'apprentissage supervisé qui permettent de discriminer efficacement des données. A l'origine, il s'agit d'un classifieur binaire destiné à discriminer des données linéairement séparables. L'objectif est de déterminer l'hyperplan optimal qui classe toutes les données en deux classes distinctes. Dans un premier temps il détermine l'ensemble des hyperplans candidats et il choisit finalement celui qui sépare « au mieux » les données, c'est-à-dire celui qui propose la plus grande marge (distance entre les vecteurs supports et l'hyperplan). Un exemple est illustré dans la figure ci-dessous.

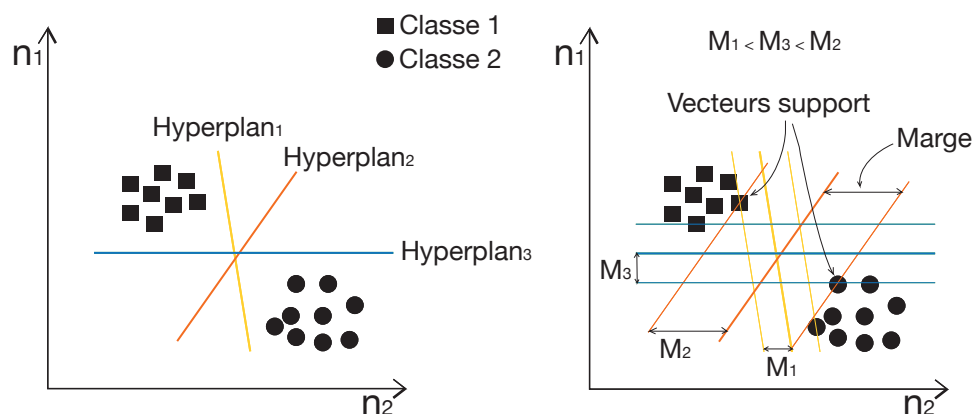


Figure 5 : Exemple d'application des SVM

Lorsque les classes ne sont pas séparables linéairement, la méthode consiste à projeter les données dans des dimensions supérieures dans lesquelles les classes

sont séparables. Pour cela, les SVM utilisent la fonction noyaux créant un espace de dimension supérieure où les données sont séparées linéairement.

Les principaux paramètres à optimiser dans l'algorithme des SVM sont :

- C le paramètre de régularisation qui permet de maximiser la marge et de réduire les échantillons d'apprentissage mal classés lors de la phase d'entraînement du modèle.
- γ qui fixe la capacité du modèle à s'adapter à un jeu d'apprentissage complexe, il équivaut à la flexibilité du modèle.

Nous utilisons ces deux algorithmes dans la suite de l'étude : les classes végétation naturelle et surfaces agricole ont été séparées grâce à l'algorithme des Forêts aléatoires (RF), et les classes prairie, maraichage et diversification, verger et canne à sucre grâce aux séparateurs à vaste marge (SVM).

Définition de la typologie des classes

Nous avons gardé la typologie des classes établie lors de la précédente étude (Bellon, 2014) à laquelle nous avons ajouté la classe Surfaces en eau, cf. tableau ci-après. Elle se base sur des observations terrain ainsi que sur la typologie utilisée par la DAAF (Direction de l'Alimentation, de l'Agriculture et de la Forêt de la Réunion) pour la Base d'Occupation du Sol (BOS). Elle se compose de 9 classes au total : cinq classes concernent les zones artificialisées et naturelles génériques, les quatre autres correspondent aux surfaces agricoles. Plus d'informations sur la typologie des classes sont disponibles en ANNEXE 1.

Tableau 1 : Typologie des classes

		Classes
1	Surfaces agricoles en culture	1.1 Canne à sucre 1.2 Prairie 1.3 Maraichage et diversification 1.4 Verger
2	Autres surfaces	2.1 Végétation naturelle 2.2 Sol Nu 2.3 Surface urbaine 2.4 Réseau routier 2.5 Surfaces en eau

II. MATÉRIEL ET MÉTHODE

IMAGES SATELLITES ET DONNÉES GÉOGRAPHIQUES

Données Raster

Les satellites optiques fournissant des images THRS sont de plus en plus nombreux. Le choix s'est porté sur la constellation Pléiades qui fournit une couverture intégrale du territoire français avec un accès simplifié aux données pour les institutions publiques. Ces deux satellites identiques, Pléiades 1A et 1B, ont été lancés à un an d'intervalle en 2011 et 2012 et opèrent sur la même orbite à 180° degrés l'un de l'autre leur permettant une capacité de revisite quotidienne.

Le programme Pléiades est issu des accords de coopération ORFEO (Optical and Radar Federated Earth Observation) mis en place en 2001 par les agences spatiales française et italienne pour développer un système d'observation dual de la Terre à résolution submétrique. Dans la continuité de ce programme ORFEO la Recette Thématique Utilisateurs (RTU) a été mise en place pour encourager les institutions et les organismes scientifiques à utiliser ces données. Ainsi, du lancement du satellite 1A jusqu'à mi-2014, les utilisateurs institutionnels bénéficiaient d'un accès gratuit aux données.

Depuis 2014, le projet EQUIPEX-GEOSUD qui vise à développer une infrastructure nationale de données satellitaires accessibles gratuitement par les utilisateurs publics, diffuse à son tour la couverture Pléiades en partenariat avec l'IGN. En effet, les images sont orthorectifiées par les services de l'IGN (« orthos-satellites ») et diffusées sur le site internet de GEOSUD. Ainsi toutes les images récupérées sont préalablement orthorectifiées et projetées en UTM 40S.

Pour chaque prise de vue, les modes panchromatique et multispectrale sont acquis. Le produit commandé peut être du type « Bundle » (panchromatique et multispectrale séparés, commande d'image effectuée en 2014) ou de type fusionné (commande effectuée en 2015). Les images sont codées en 16bits dont seuls 12 sont significatifs. Les images panchromatiques sont acquises à la résolution de 0.7m, les multispectrales à 2.8m et sont ensuite ré-échantillonnées respectivement à 0.5m et 2m.

Une couverture complète de l'île est disponible. Seule la zone de Saint-Pierre est couverte par plusieurs images prises entre 2012 et 2014 : au total 5 images couvrent cette zone. Au cours de la précédente étude réalisée en 2014 sur Saint-Pierre (Bellon, 2014), trois images Pléiades ont été traitées celles du 27/07/2012, 28/05/2013 et 04/01/2014. La méthode de classification de 2014 a été développée sur l'image du 27/07/2012. En supplément, nous avons récupéré deux images supplémentaires couvrant la zone de Saint-Pierre : celles du 04/03/2013 et du 02/09/2013.

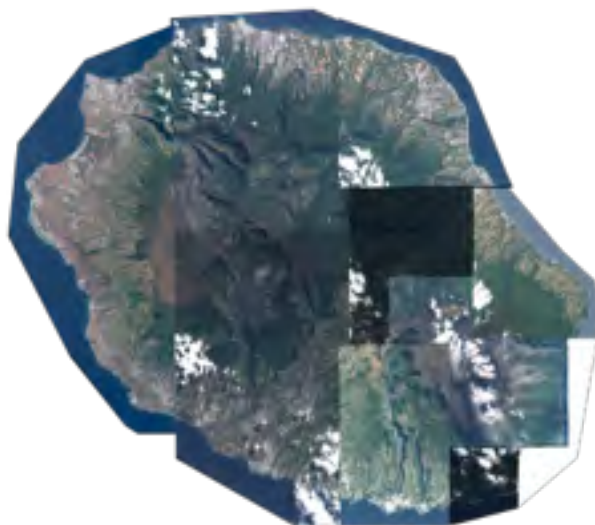


Figure 6 : Aperçu des images Pléiades couvrant l'ensemble du territoire

Les caractéristiques de chaque image utilisée pour l'étude sont présentées dans l'ANNEXE 2.

En complément des images satellites, nous avons également utilisé le MNT Litto 3D à 5m de résolution, produit et diffusé par l'IGN.

Données Vecteur

Une série de données vectorielles est utilisée pour faciliter l'extraction de certaines classes d'occupation du sol, notamment le réseau routier, le réseau hydrographique et les surfaces en eau. Nous avons également utilisé les bases de données de l'occupation du sol de la Réunion (BOS) 2013 et 2014 pour nous aider à établir nos différents jeux de données d'apprentissage. Les réseaux hydrographique, surfaces en eau et réseau routier sont issus de la BDTOPO® 2015 de l'IGN. Quant à la BOS, elle est produite par la Direction de l'Alimentation, de l'Agriculture et de la Forêt de la Réunion (DAAF) d'après les déclarations des agriculteurs.

PRÉPARATION DES DONNÉES

Images Pléiades

Les images Pléiades ont été traitées suivant la méthodologie mise en place dans l'étude réalisée en 2014 : une première phase de préparation des images (corrections atmosphériques, fusion,...) et une deuxième phase de calcul d'une bande texturale.

La conversion des comptes numériques en réflectance TOA est nécessaire car celle-ci permet de récupérer les valeurs physiques des propriétés des surfaces de chaque pixel. On calcule tout d'abord la luminance en entrée du capteur suivant la relation suivante :

$$L^{\lambda}_{TOA} = X^{\lambda} / a_{\lambda}$$

L^{λ}_{TOA} : luminance : puissance émise dans une direction donnée par unité de surface et par unité d'angle solide (en $W.m^{-2}.sr^{-1}.\mu m^{-1}$).

X^{λ} : valeur en compte numérique

a_{λ} : produit du coefficient d'étalonnage absolu et du coefficient de sensibilité inter-détecteurs (cf. gain dans les métadonnées).

Les luminances de chaque bande sont ensuite transformées en réflectance top of atmosphere (TOA) ce qui permet de normaliser les paramètres d'éclairement en prenant en compte les paramètres du flux solaire incident au sommet de l'atmosphère selon la formule :

$$\rho^{\lambda}_{TOA} = [\pi . L^{\lambda}_{TOA}] / [E^{\lambda}_s . \cos \theta_s . (d/d_0)]$$

ρ^{λ}_{TOA} : réflectance TOA, rapport entre l'énergie réfléchie et l'énergie solaire incidente

E^{λ}_s : éclairement solaire moyen hors atmosphère (*Solar Irradiance* - métadonnées)

θ_s : angle zénithal solaire en degrés (*Sun élévation* - métadonnées)

d/d_0 ratio entre la distance Terre-Soleil à la date d'acquisition et la distance Terre-Soleil moyenne (en jour julien).

Les valeurs physiques en réflectance sont ensuite multipliées par 1000 pour faciliter les traitements postérieurs. Ce calcul a été réalisé avec le logiciel ERDAS Imagine (modèle de calcul établi grâce au module Model Maker).

La fusion des images multispectrale et panchromatique permet de faire bénéficier à l'image MS de la haute résolution spatiale et de la forte texture de l'image PAN. Cette technique modifie légèrement les valeurs radiométriques de l'image multispectrale originale mais elle est tout de même fortement conseillée, particulièrement pour les analyser par approche orientée objet (Maxwell, 2005), elle permet une analyse plus complète et plus fine des scènes observées.

On distingue deux pré-traitements différents suivant le format de livraison des images dans la préparation de ces dernières :

- Images de type bundle : dans un premier temps une correction atmosphérique est appliquée aux images multispectrales (MS) et panchromatiques (PAN) de manière à passer des comptes numériques en réflectance Top Of Atmosphere (TOA). Dans un deuxième temps, les images MS et PAN sont fusionnées (module pansharpening du logiciel Monteverdi) et enfin, une série de masques est appliquée (océan, nuages et pixels saturés). Les différentes étapes de traitement sont présentées dans la figure ci-après.
- Images de type fusionné : ces images sont livrées en dalle de 1km² les traitements de correction et de fusion étant déjà faits, nous les avons rassemblées (mosaïquage) pour ensuite appliquer nos différents masques. Les étapes sont présentées dans l'ANNEXE 3.

Une bande texturale a été calculée afin de discriminer plus efficacement les surfaces urbaines. En effet, les surfaces urbaines étant fortement hétérogènes, elles bénéficient d'une texture très marquée. Ainsi, nous avons calculé, à partir de la bande verte, une bande texturale basée sur la distance moyenne euclidienne dans une fenêtre de 50 x 50 pixels.

MNT Litto 3D et extraction des zones de fortes pentes

Le relief est une des principales contraintes qui limite l'aménagement et le développement de l'île. Il est donc utile d'insérer des informations relatives aux pentes dans notre classification. Ces informations nous servent à classer la végétation naturelle plus aisément.

A partir du MNT, nous calculons l'ensemble des pentes que nous reclassons ensuite en deux classes : les pentes supérieures et les pentes inférieures à 30 degrés. Les

zones où les pentes sont trop fortes sont très difficilement cultivables puisqu'inaccessibles par les engins agricoles. Ce sont donc des zones où la végétation naturelle domine. Même s'il existe des zones cultivées sur des pentes supérieures à 30° principalement aux pieds des ravines, celles-ci sont négligeables et difficilement repérables sur des images satellites. Nous les considérerons comme de la végétation naturelle dans notre classification.

Ayant un grand nombre d'images à traiter nous avons établi un modèle avec le module ModelBuilder du logiciel Arcgis pour faciliter l'extraction des données sur l'emprise des images satellites. Ce modèle est présenté dans la figure ci-dessous. En sortie nous obtenons un fichier vectoriel contenant deux entités : une première portant le code 1, elle correspond aux zones où les pentes sont inférieures à 30 degrés, l'autre porte le code 2 et correspond au reste de l'île (pentes > 30°).

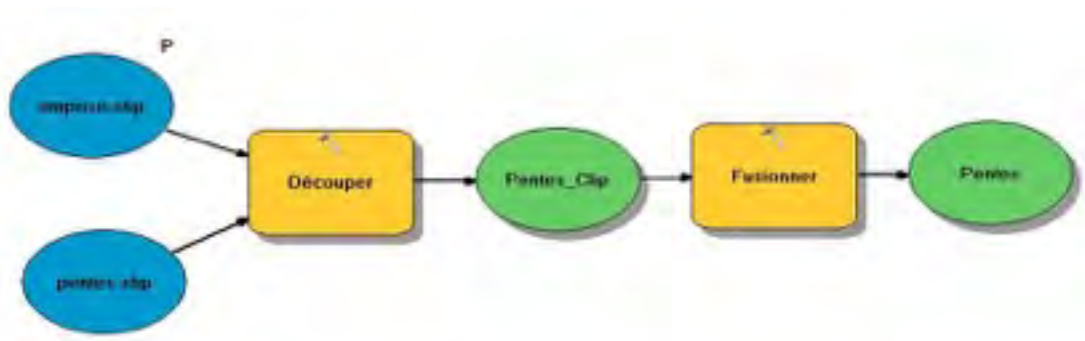


Figure 7 : Modèle Builder permettant l'extraction des pentes sur les zones étudiées

Données vecteur

Nous utilisons les couches vecteurs Route et Hydrographie comme couche thématique dans notre segmentation pour contraindre les limites de celle-ci et ainsi classer rapidement les objets appartenant à ces classes. Pour le réseau routier nous avons, comme dans l'étude de 2014 (Bellon, 2014), calculé des zones tampon de 5m pour les routes à une voie et 7m pour les routes à deux voies et plus. Pour le réseau hydrographique nous avons calculé des zones tampon de 3m autour des bras de rivières secondaires et de 5m autour des lits principaux. Nous l'avons ensuite joint à la couche Surface en eau qui recense les bassins et les réservoirs du territoire.

Nous avons aussi établi deux modèles avec le module ModelBuilder pour extraire les données du réseau routier et des surfaces en eau suivant les emprises des images étudiées.

BASE DE DONNÉES

Variables descriptives

Afin de caractériser chaque objet créé lors des segmentations, nous avons calculé au total 31 variables descriptives : 23 indices radiométriques et 8 indices texturaux. Les indices radiométriques rassemblent :

- les moyennes des valeurs radiométriques des pixels composant un objet dans une bande spectrale donnée,
- les écart-types des valeurs radiométriques des pixels dans une bande donnée,
- la somme des réflectances qui calcule la moyenne des valeurs de brillance pour chaque bande spectrale,
- divers indices caractéristiques de la végétation tels que le NDVI (Tucker, 1979), l'indice de végétation brut (Richardson et Wiegand (1977), l'indice de couleur de la végétation, etc.
- la moyenne de la distance euclidienne entre la valeur de chaque pixel dans une fenêtre donnée (50x50 pixels) et la valeur du pixel central dans cette fenêtre pour une bande données (bande verte dans notre cas),

Les indices texturaux développés par Haralick (1979), qui permettent de calculer un grand nombre de paramètres statistiques décrivant différents aspects de la texture tels que le contraste, l'homogénéité, l'entropie, etc. Chaque indice a été calculé dans toutes les directions et pour toutes les bandes mobilisées (B, V, R et PIR).

De plus amples informations sur les indices sont disponibles en ANNEXE 3.

Jeu de données

Pour identifier et analyser au mieux les différents types de cultures étudiés nous avons établi une base d'apprentissage pour chaque image couvrant la zone de Saint-Pierre. Une première base (*jeu de données 1*) identifie des échantillons de végétation naturelle et de surfaces agricoles, une deuxième (*jeu de données 2*) identifie des parcelles de canne à sucre, prairie, verger et maraichage et diversification.

Ces bases d'apprentissage reprennent les échantillons utilisés dans l'étude de 2014 réalisées à partir de l'image du 27/07/2012. Nous vérifions par photo-interprétation si la nature des individus a changé entre 2012, 2013 et 2014, nous supprimons les individus qui se retrouvent sous les nuages, modifions le type des individus lorsque l'on observe un changement et nous complétons le jeu par le nombre d'individus supprimés toujours par photo-interprétation. De plus amples informations sur le contenu des jeux de données sont disponibles en ANNEXE 4.

ADAPTATION DU MODÈLE DE 2014

L'adaptation du modèle de classification fait en 2014 s'articule en deux temps : tout d'abord l'adaptation à d'autres dates sur la zone de Saint Pierre, puis sur le reste du territoire. Il s'agit ici d'évaluer les paramètres de segmentation et de vérifier leur efficacité sur d'autres images et ensuite de tester les méthodes de classification et de déterminer celles qui proposent les meilleurs résultats pour ensuite les appliquer à l'ensemble des images disponibles sur le territoire. L'extraction des classes d'occupation et d'utilisation du sol s'articule en trois temps : tout d'abord l'extraction des classes générales d'occupation du sol : réseau routier, surfaces urbaines, surfaces en eau, sol nu et végétation, la discrimination des classes de végétation naturelle et surfaces agricoles et enfin la classification des classes culturelles.

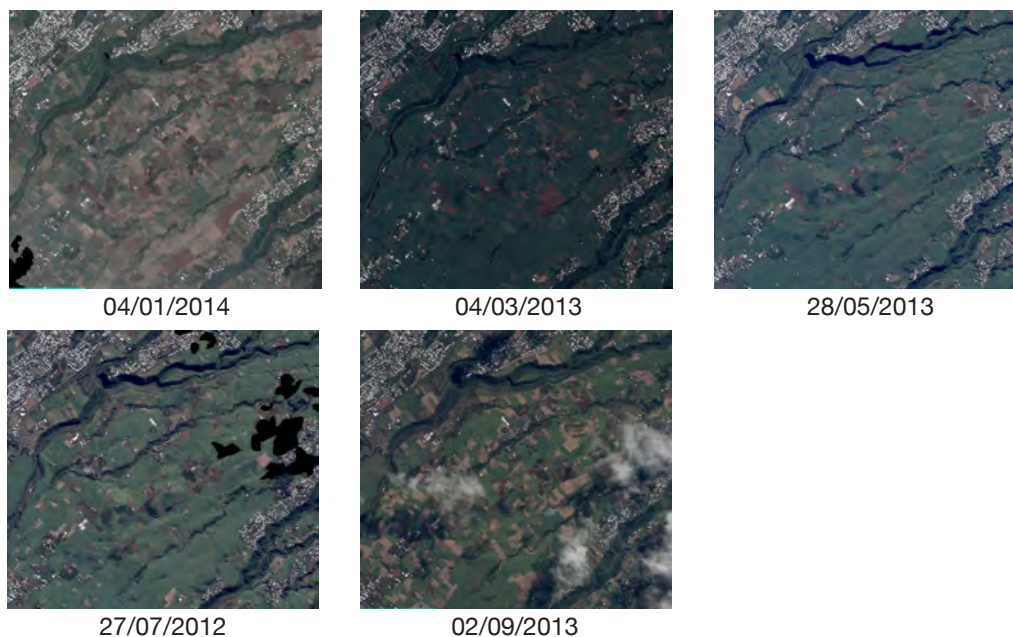


Figure 8 : Aperçu des images satellites couvrant la zone de Saint-Pierre

La méthode de classification de 2014 a été établie sur l'image du 27/07/2012, les autres images couvrant la zone de Saint-Pierre se répartissent tout au long de l'année : 03/04/2013, 28/05/2013, 02/09/2013 et 04/01/2014. Les cultures sont donc à des niveaux de croissance différents sur chacune des images ce qui peut influencer fortement leurs réponses spectrales. Le changement le plus visible s'observe sur la canne à sucre et particulièrement entre les images où la campagne sucrière n'a pas encore eu lieu et celles où les cannes sont coupées (la campagne sucrière débute mi-juillet et finit début décembre).

Segmentation des images

La méthodologie réalisée en 2014 proposait quatre segmentations multirésolution successives pour classer l'image satellite et discriminer au total 8 classes d'occupation du sol. Afin d'optimiser le modèle et surtout de diminuer les temps des traitements sous eCognition, nous avons réduit le nombre de segmentation à trois et nous introduisons des informations relatives aux surfaces en eau et aux pentes pour minimiser les erreurs de classification d'objets.

Les critères de segmentation définis en 2014 ne permettait de classer les zones en ombres sur l'image de 2012. Or, nous observons que ces zones sont situées sur les remparts des ravines et les seuils établis ne sont pas efficaces pour discriminer ces zones sous ombre sur les autres images. Nous introduisons donc des informations relatives aux pentes classer ces zones directement en surface à agricole.

Les procédures de segmentation sont constituées de trois étapes (tableau 4) qui sont appliquées successivement aux images. Les seuils retenus sont ceux définis dans l'étude de 2014, ils s'adaptent bien aux nouvelles images et délimitent efficacement les entités étudiées.

Tableau 2 : Récapitulatif des paramètres de segmentation utilisés

Algorithme	Informations supplémentaires	Echelle	Couleur	Forme
1 - Multirésolution	Routes, surfaces en eau, pentes	1000	0.7	1
2 - Multirésolution	-	30	0.9	1
3 - Multirésolution	-	100	0.8	1

Chaque niveau de segmentation permet de délimiter des objets appartenant à une ou plusieurs classes d'occupation du sol : nous extrayons en premier les objets de grandes tailles pour affiner progressivement et arriver à la segmentation d'objets plus petits.

- un premier niveau de segmentation utilise les quatre bandes de l'image satellite (B, V, R, et PIR), la bande texturale et les couches thématiques routes, pentes et hydrographie. Cette segmentation génère des objets assez grands (à l'échelle des quartiers) et suit les limites des couches thématiques.
- le deuxième niveau de segmentation suit les limites de la précédente, et crée des objets de petites tailles adaptés à la détection des bâtis individuels.
- un troisième niveau de segmentation est créé indépendamment des deux autres (les objets non classés issus des deux premières segmentation sont fusionnés), il génère des objets de tailles équivalentes aux parcelles agricoles.

Les paramètres de segmentation sont délicats à mettre en place puisqu'aucune technique n'existe pour les déterminer automatiquement, l'opérateur doit donc tester différentes valeurs jusqu'à trouver celles qui délimitent le mieux aux objets observés. La validation de ces derniers se fait par interprétation visuelle.

Les opérations de segmentation ont été réalisées avec le logiciel eCognition version 8.0 (Trimble(R)).

Classification des objets

Extraction des classes générales d'occupation du sol par règles expertes

Les classes générales d'occupation du sol (tissu urbain, réseau routier, surface en eau, surface bâtie, sol nu et végétation) sont classées par règles expertes, c'est à dire que l'opérateur sélectionne un nombre réduit d'indices caractéristiques des objets étudiés et fixe des seuils pour chaque classe. Cette classification utilise la logique booléenne : l'objet appartient ou n'appartient pas à la classe définie.

Nous gardons les mêmes indices discriminants utilisés dans l'étude de 2014 et nous ajustons les seuils par interprétation visuelle non pas sur une seule image (27/07/2012) mais sur les quatre nouvelles images disponibles sur la zone de Saint-Pierre (05/28/2013, 03/04/2013, 02/09/2013 et 04/01/2014).

Tableau 3 : Seuils retenus pour extraire les classes d'occupation du sol par règles expertes

	Indices	27 Juillet 2012 (ETUDE 2014)	28 Mai 2013	3 Avril 2013	4 Janvier 2014	2 Septembre 2013
Réseau routier	Layer thématique (shp) - Routes	«ID» = 1	«ID» = 1	«ID» = 1	«ID» = 1	«ID» = 1
Tissu urbain	Mean Dist Euclidienne	>= 255	>= 220	>= 220	>= 220	>= 220
Végétation sous ombre	Somme des réflectances NDVI	<= 600 <= 0.28	-	-	-	-
Surface en eau	Layer thématique (shp) - Hydro	-	«ID» = 1	«ID» = 1	«ID» = 1	«ID» = 1
Végétation ravines	layer thématique (shp) - Pentes	-	«ID» = 2	«ID» = 2	«ID» = 2	«ID» = 2
Surface bâtie	Max diff Mean Dist Eucli Mean Rouge Mean Vert NDVI Rapport \sum refl / PIR	<=1.48 >= 259 >= 1007 >= 830 <= 0.41 >= 2.67	<=1.1 >= 220 >= 900 >= 580 <= 0.41 >= 2.5	<=1.1 >= 220 >= 900 >= 580 <= 0.41 >= 2.5	<=1.1 >= 220 >= 900 >= 580 <= 0.41 >= 2.5	<= 1.1 >= 220 >= 900 >= 580 <= 0.41 >= 2.5
Sol nu	NDVI Rapport \sum refl / PIR	<= 0.3 >= 2.25	<= 0.3 >= 2.25	<= 0.3 >= 2.25	<= 0.3 >= 2.25	<= 0.3 >= 2.25
Végétation						

Extraction des classes de végétation par algorithme de fouille de données

Classification des classes végétation naturelle et Surfaces agricoles en culture

Une fois les classes générales d'occupation du sol extraites, les objets restants, non classés, forment la classe végétation. Nous avons tout d'abord exporté en fichier vectoriel l'ensemble des informations relatives à ces objets (informations spectrales, texturale, ... soit un total de 31 variables descriptives).

Plusieurs algorithmes, parmi ceux proposés par le logiciel Weka 3.6 ont été testés sur chaque jeux de données établis pour chaque image. L'algorithme Random Forest offre les meilleurs résultats pour discrétiser les classes «végétation naturelle» et «surface agricole».

L'optimisation des paramètres est faite grâce à l'outil CVPParameterSelection disponible dans Weka qui permet de tester différentes valeurs de paramètres et de déterminer les plus efficaces. Ainsi, le meilleur nombre d'arbres, entre 100 et 1000, est de 400 et le nombre de variables à prendre en considération dans chaque arbre est de 6.

La méthode de classification choisie pour différencier les classes Végétation naturelle et Surfaces agricoles en culture dans l'étude de 2014 est une classification supervisée basée sur l'algorithme des Maximum de vraisemblance et sur 5 variables d'importance. Ces cinq variables discriminantes avaient été déterminées grâce à l'algorithme des forêts aléatoires qui permet, entre autre, d'extraire d'une base de données les variables descriptives qui caractérisent le plus les entités étudiées. Or, nous avons vu dans la littérature scientifique que cet algorithme offre de moins bonnes performances et qu'il n'est pas aussi fiable que les algorithmes de fouille de données sur des base de données importante avec un grand nombre de variables descriptives. L'utilisation de l'algorithme des forêts aléatoires nous permet de n'omettre aucune variables et de toutes les utilisées dans les construction d'arbres. De plus, la méthode de classification supervisée utilisée dans l'étude de 2014 ne permet pas de reproductibilité du modèle à l'inverse de l'algorithme des forêts aléatoires qui génère un modèle de classification ré-utilisable et applicable sur différents jeux de données. Après analyse des bases d'apprentissage il s'avère que les 5 variables d'importance non-corrélées déterminées par l'algorithme des forêts aléatoires ne sont pas tout à fait les mêmes pour chaque date.

Classification des classes canne à sucre, verger, prairie, maraichage et diversification

Pour la discrimination des parcelles agricoles en culture notre choix s'est porté sur une classification utilisant l'algorithme des séparateurs à vaste marge, parmi les algorithmes testés il est celui qui offre les meilleurs résultats de classification pour la discrimination des classes canne à sucre, verger, prairie et maraichage et diversification.

L'algorithme des séparateurs à vaste marge a été implémenté dans Weka par John Platt (1998) sous le nom de SMO, il gère les classifications multi-classes en utilisant la séparation binaire. Plusieurs paramètres peuvent être modifiés pour optimiser l'algorithme : C le paramètre de régularisation, γ il équivaut à la flexibilité du modèle et le type de noyau à utiliser pour la projection des données dans des dimensions supérieures. Nous avons également utilisé la fonction CVParmetreSelection de Weka pour optimiser les paramètres des SVM. Les valeurs les plus efficaces pour les paramètres sont de $C = 1$, $\gamma = 1.0E^{-12}$, le noyau à utiliser et le RBF Kernel.

Dans l'étude de 2014, les parcelles agricoles en culture, les objets ont été classés grâce à une analyse en composantes principales (ACP) : la combinaison des indices en trois axes synthétiques maximiserait les possibilités de discrimination tout en réduisant le volume des variables et donc le temps de traitement. Le choix a été de classer les quatre types de cultures grâce aux axes issus de l'ACP en regardant les valeurs médiane, minimale et maximale pour chaque classe. Les règles de classification sont basées sur la logique floue où le domaine de définition de chaque classe est fixé par la combinaison des valeurs min, max et médiane de chaque composante principale (Bellon, 2014). Lors de nos tests de reproductibilité, il apparaît que les individus répertoriés dans nos bases d'apprentissage se confondent fortement particulièrement les entités représentatives du maraichage qui se confondent avec toutes les autres classes. De surcroît, les composantes principales obtenues pour les images du 28/05/2013 et 04/01/2014 ne synthétisent que 72% et 76% de l'information soit environ 10% de moins que dans l'étude de 2014 où 84% de l'information était synthétisée.

La synthèse des étapes de segmentation/classification des classes d'occupation et d'utilisation du sol est présentée ci-dessous (figure 9).

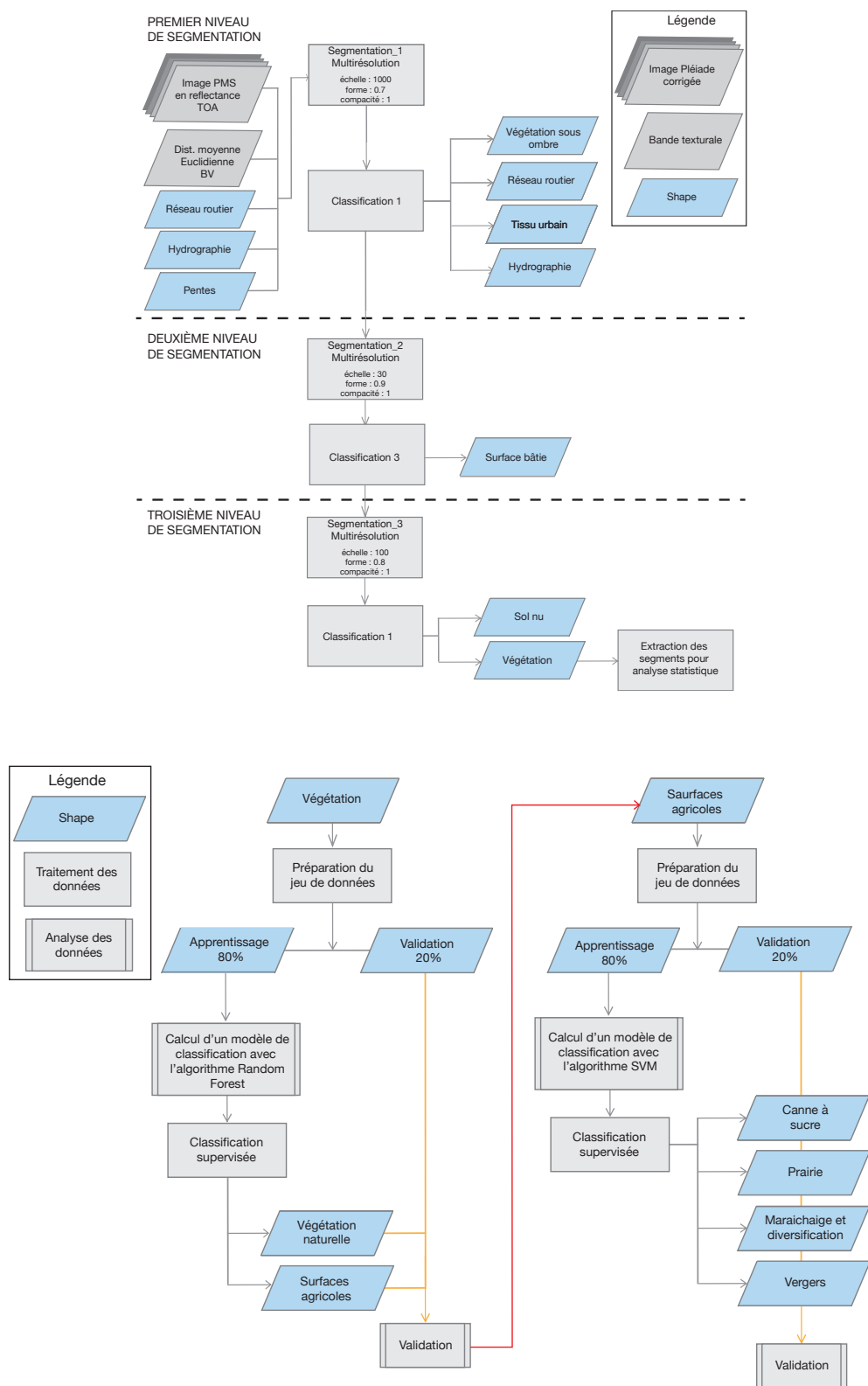


Figure 9 : Synthèse des traitements successifs de segmentation / classification

TEST DE REPRODUCTIBILITÉ SUR D'AUTRES RÉGIONS

Nous avons ensuite testé la méthode de classification modifiée issue des classifications des images de Saint Pierre (03/04/2014, 28/05/2013, 02/09/2013 et 04/01/2014). Au total, 7 images satellites couvrent le territoire insulaire. L'intégralité des données a été préparées : mosaïquage, masquage, extraction des données vectorielles sur les emprises des images.

Nous avons extrait les classes générales d'occupation du sol sur 4 d'entre elles : celles du 20/08/2013 Nord et Sud (nous l'avons découpée en deux pour permettre et alléger les traitements dans eCognition) celle-ci couvre tout l'Est de l'île. Celle du 02/09/2013 partie Nord qui couvre le centre nord de l'île et celle du 27/08/2013 qui couvre le nord ouest de l'île.



Figure 10 : Images Pléiades traitées sur le reste du territoire

Une base d'apprentissage a été créée sur l'image du 02/09/2013 Nord pour permettre de valider la classification de la végétation. Les jeux de données ont été mis en place par photo interprétation et certaines entités ont fait l'objet d'une vérification terrain.

L'évaluation de la qualité de l'ensemble des classifications a été effectuée grâce au pourcentage d'entités bien classées et à l'indice de Kappa dont la valeur est comprise entre 0 et 1. Ce coefficient tient compte des erreurs d'attribution des entités en lignes et en colonnes. Un coefficient de Kappa supérieur à 0,8 indique une très bonne qualité de classification, un coefficient compris entre 0,6 et 0,8 indique une classification de bonne qualité, en dessous de 0,6 la qualité de classification est faible.

III. ANALYSE DES RÉSULTATS

ADAPTATION DU MODÈLE SUR LA ZONE DE SAINT-PIERRE À D'AUTRES DATES

Segmentation des images

Les paramètres utilisés en 2014 pour les segmentations de l'image de 2012 permettent une bonne délimitation des entités aux autres dates (cf figure ci-dessous), ils restent appropriés pour l'analyse des autres images, nous ne modifions pas les paramètres.

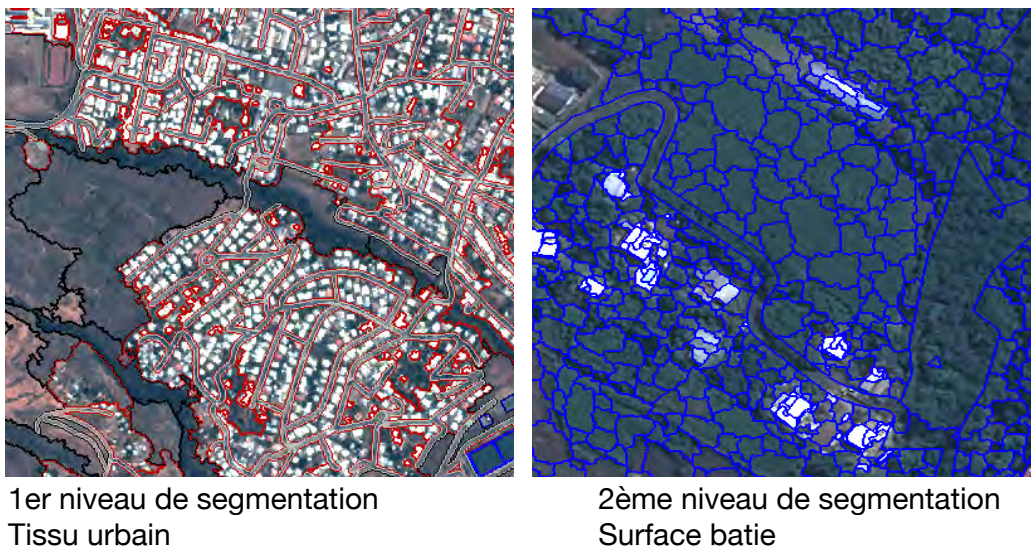


Figure 11 : Aperçu des objets issus des segmentations de niveau 1 et 2

Malgré tout, nous trouvons une alternative intéressante concernant la segmentation des objets représentatifs de la végétation sous ombre. Ces zones situées essentiellement sur les remparts des ravines sont plus rapidement identifiable à partir des informations relatives aux pentes issues du traitement du MNT.

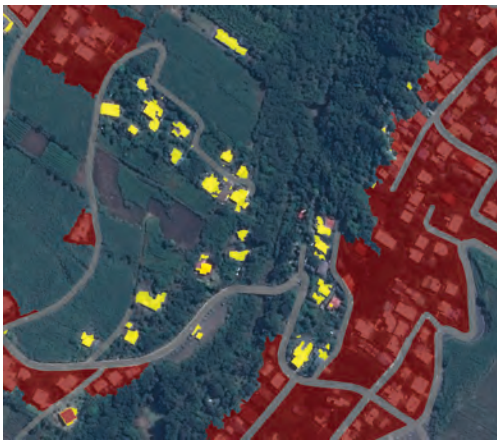
Classification des objets

Classification des classes générales d'occupation du sol

Nous observons de mauvaises attributions de classes avec les règles expertes choisies pour la classification de l'image de 2012. En effet, l'utilisation du seuil défini

dans l'étude de 2014 pour extraire le tissu urbain entraîne une sous-détection des objets, nous avons diminué le seuil progressivement jusqu'à l'obtention d'un seuil efficace sur chaque image.

Nous observons aussi quelques mauvaises détections de bâti isolé dans le fond des ravines (lits de rivières), les caractéristiques radiométriques du bâti individuel étant proche de celui des sols nu. Pour éviter ces confusions entre le sol nu des fonds de ravines et les surfaces bâties, nous modifions légèrement les seuils et ajoutons un autre layer thématique des surfaces en eau qui permet de classer rapidement les fonds de ravines, les lits de rivières, les bassins et les réservoirs éliminant les mauvaise détection de bâti individuel.



Classification de la surface bâtie



Classification du tissu urbain

Figure 12 : Aperçu des classification du Tissu urbain et des Surfaces bâties après ajustement des seuils

Les sols nus sont également source de confusion, particulièrement sur les images du 04/01/2014 et du 02/09/2013 : les parcelles agricoles où la canne à sucre est coupée sont classées en sol nu, aucune modification de seuils n'a permis de les exclure de cette classe. Nous avons donc une sous-détection des parcelles agricoles à ces dates là.

Nous validons par photo-interprétation la classification des 5 classes générales d'occupation du sol, à savoir le tissu urbain, le réseau routier, les surfaces en eau, les sols nu et la végétation. L'arbre des processus de traitement «process tree» eCognition sera appliqué sur chacune des images étudiées.

Extraction des classes d'utilisation du sol par algorithme de fouille de données

Classification des classes Végétation naturelle et Surfaces agricoles en culture

La discrimination des classes Végétation naturelle et Surface agricole grâce avec l'algorithme des forêts aléatoire obtient de bons résultats avec un indice de kappa supérieur à 0.8 : les modèles établis à partir des bases d'apprentissage des images du 03/04/2013, 28/05/2013, et 04/01/2014, appliqués sur ces mêmes images atteignent respectivement un coefficient de Kappa de 0.9, 0.8238 et 0.8506. Les résultats relatifs à l'image du 02/09/2013 sont légèrement inférieurs mais restent bons avec un coefficient de Kappa égale à 0.788 (cf figure ci-dessous). Nous observons visuellement que les principales sources d'erreurs proviennent de confusions dans la classe Verger où des segments se retrouvent classés en végétation naturelle du fait de leur caractéristiques très proches.

En comparaison à la méthode de classification utilisée dans l'étude de 2014 (classification supervisée par Maximum de vraisemblance) , les résultats obtenus avec l'algorithme des forêts aléatoires sont meilleurs (cf, figure 11 ci-dessous).

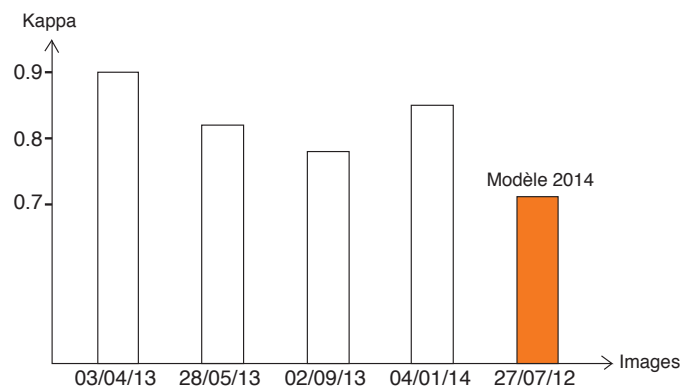


Figure 13 : Comparaison des coefficient de Kappa obtenus pour les classifications avec l'algorithme des Random Forest pour l'ensemble des images couvrant la zone de Saint Pierre

L'indice de Kappa obtenu avec le modèle de 2014 est de 0,73. Nous obtenons donc des modèles de classification plus performants que celui établi en 2014, les différentes classes d'utilisation du sol sont mieux discriminées et la méthode de classification choisie permet une reproductibilité du modèle sur différents jeux de données.

En revanche, les tests de transfert d'un modèle établi à une date sur une autre date ne sont pas très concluants, nous obtenons entre 50 et 76% d'entités bien classées ce qui équivaut à des indices de Kappa oscillant entre 0,0112 et 0,5295 cf tableaux ci-dessous.

Tableau 4 : Test de transfert des modèles établis à partir de chaque image

Modèles/ Images	28 Mai 2013	4 Janvier 2014	3 Avril 2013	2 Septembre 2013
28 Mai 2013	91.3% Kappa = 0.8238	54.6 % Kappa = 0.0601	64.05% Kappa = 0.354	44.32% Kappa = 0,0453
4 Janvier 2014	54.09% Kappa = 0.0112	92.5% Kappa = 0.8506	63.48% Kappa = 0.0266	57.17% Kappa = 0,0114
3 Avril 2013	55,45% Kappa = 0.0697	54.5 % Kappa = 0.0557	95.55% Kappa = 0.9	75.8% Kappa = 0,5295
2 Septembre 2013	53,63% Kappa = 0.0543	54.6 % Kappa = 0.0624	76.40% Kappa = 0.4183	89.7% Kappa = 0.788

Les indices de Kappa inférieurs à 0,6 reflètent une classification de qualité très moyenne, un trop grand nombre d'entités sont mal classées. Nous ne pouvons donc valider un modèle unique qui permettrait de classer l'ensemble des classes d'utilisation du sol. Il faudra choisir le modèle dont la date d'élaboration se rapproche le plus de la date de prise de vue de l'image analysée.

Classification des cultures

La classification des quatre classes culturelles basée sur l'algorithme des séparateurs à vaste marge atteint des précisions très satisfaisantes : le modèle spécifique établi à partir de l'image du 04/01/2014 offre les meilleurs résultats sur cette même image avec un coefficient de Kappa de 0.8944 et 95% d'entités bien classées, les autres modèles sont légèrement moins performants avec un coefficient de Kappa inférieur à 0.8 : 0.7673 pour l'image du 28/05/2013, 0.71 pour l'image du 02/09/2013 et 0.6425 pour l'image du 03/04/2013 cf. figure ci-dessous.

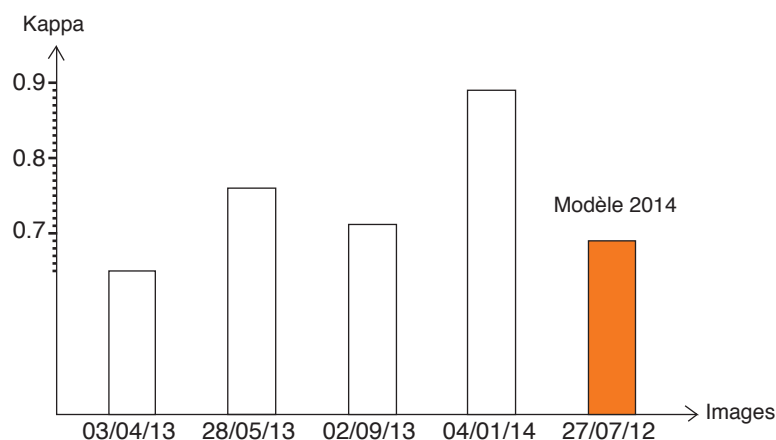


Figure 14 : Comparaison des coefficients de Kappa obtenus pour les modèles de classification élaborés avec l'algorithme des SVM

En comparaison à la méthode de 2014, les résultats sont meilleurs pour les modèles établis à partir des images du 28/05 et 04/01. En effet, le kappa obtenu avec le modèle de 2014 (classification supervisée à partir des trois premières composantes principales et règles basées sur la logique floue) est de 0.70. Le modèle établi à partir de l'images du 02/09 fournit des performances équivalentes alors que le modèle établi à partir de l'image du 03/04 est bien moins performant.

TEST DE REPRODUCTIBILITÉ SUR D'AUTRES RÉGIONS

Segmentation des image

Les différents paramètres fixés pour les segmentations s'adaptent correctement aux images couvrant les autres régions de l'île, la délimitation des objets reste appropriée aux entités étudiées. Nous gardons donc les valeurs choisies lors de l'étude de 2014 (cf tableau partie matériel et méthode).

Classification des objets

Classification des classes générales d'occupation du sol

Les seuils retenus pour discriminer les classes générales d'occupation nécessitent des modifications, ils ne sont pas efficaces sur toutes les images analysées. L'harmonisation des seuils pour la classification par règles expertes est une tâche longue et délicate. Les valeurs qui produisent les meilleurs résultats de classification sont détaillés dans le tableau ci-après.

Tableau 5 : seuils de classification utilisés pour l'extraction des classes générales d'occupation du sol

Classe / Image	Indices	02/09/13 N	27/08/13	20/08/13 S	20/08/13 N
Réseau routier	Layer thématique (shp) Routes	«ID» =1	«ID» =1	«ID» =1	«ID» =1
Tissu urbain	Mean Dist Euclidienne	>= 200	>= 200	>= 200	>= 200
Végétation sous ombre	Somme des réflectances NDVI	-	-	-	-
Surface en eau	Layer thématique (shp) Hydrographie	«ID» = 1	«ID» = 1	«ID» = 1	«ID» = 1
Végétation ravines	layer thématique (shp) pentes	«ID»= 2	«ID»= 2	«ID»= 2	«ID»= 2
Surface bâtie	Max diff	<= 1.4	<= 1.4	<= 1.4	<= 1.4
	Mean Dist Eucli	>= 180	>= 180	>= 180	>= 180
	Mean Rouge	>= 600	>= 600	>= 600	>= 600
	Mean Vert	>= 580	>= 580	>= 580	>= 580
	NDVI	<= 0.41	<= 0.41	<= 0.41	<= 0.41
	Rapport $\sum \text{refl} / \text{PIR}$	>= 2.33	>= 2.33	>= 2.33	>= 2.33
Sol nu	NDVI				
	Rapport $\sum \text{refl} / \text{PIR}$				
Végétation		unclassified	unclassified	unclassified	unclassified

Extraction des classes d'utilisation du sol par algorithme de fouille de données

Classification des classes Végétation naturelle et Surfaces agricoles en culture

L'application du modèle de classification par algorithme des Random Forest établi sur la zone de Saint-Pierre à partir du secteur sud de l'image du 02/09/2013 sur la partie nord (commune de Saint-Denis, Sainte Suzanne) fournit un taux de réussite de 85.8% et un Kappa de 0.71.

Tableau 6 : Matrice de confusion et précision globale de la classification par l'algorithme des Random Forest sur la zone Nord de l'image du 02/09/2013

		Objets de référence				
		Végétation naturelle	Surface agricole	Total	Précision producteur	Précision utilisateur
Objets classés	Végétation naturelle	272	80	352	89.8%	77.3%
	Surface agricole	31	397	428	83.2%	92.7
	Total	303	477	780	Précision globale = 85.8% Kappa = 0.71	

Ce résultat est satisfaisant; le modèle fonctionne bien sur la partie nord de l'image. La végétation naturelle et les surfaces agricoles sont bien discriminées.

Classification des cultures

Nous appliquons le modèle de classification basé sur les SVM créé à partir de l'image du 02/09/2013 (secteur sud) sur le secteur nord. La matrice de confusion obtenue sur un jeu de validation de 478 objets est présentée dans le tableau ci-dessous.

Tableau 7 : Matrice de confusion et précision globale de la classification par l'algorithme des Random Forest sur la zone Nord de l'image du 02/09/2013

		Objets de référence						
		Canne à sucre	Verger	Marai-chage	Prairie	Total	Précision producteur	Précision utilisateur
Objets classés	Canne à sucre	114	6	16	9	145	65.1%	78.6%
	Verger	14	110	16	4	144	70.5%	76.4%
	Marai-chage	32	39	54	1	126	60.7%	42.8%
	Prairie	15	1	3	44	63	75.9%	69.9%
	Total	175	156	89	58	478	précision globale = 67.3% Kappa = 0.55	

Le transfert de ce modèle à d'autres région de l'île fournit des résultats juste corrects avec une précision globale de 67% pour discriminer les cultures. Il reste à tester ce modèle à d'autres régions de l'île pour valider la méthodologie complète.

IV. DISCUSSION

LIMITES ET PERSPECTIVES

L'objectif de la méthode présentée ici est l'adaptation et l'amélioration d'une méthode de cartographie de l'occupation et de l'utilisation du sol par analyse orientée objet d'images satellites à très haute résolution spatiale. Les méthodes de classification choisies pour discriminer les classes de végétation fournissent des résultats satisfaisants et permettent la reproductibilité des modèles. Mais pour que cette méthode soit vraiment efficace sur d'autres zones d'études il faudra privilégier l'analyse d'images prises à des dates proches de celles utilisées dans l'étude présentée ici.

L'analyse orientée objet d'images à très haute résolution spatiale a montré son efficacité et son potentiel pour cartographier l'occupation et l'utilisation du sol cependant, ce genre d'étude induit l'utilisation de ressources informatiques très importantes et reste très chronophage. Mettre en place une méthode de classification de l'occupation et de l'utilisation du sol opérationnelle sur l'ensemble du territoire est un véritable défi. L'élaboration de telles cartes sera difficilement automatisable et nécessitera l'intervention d'un opérateur qualifié. L'élaboration d'un modèle unique performant à n'importe quelle date et pour toutes les régions de l'île apparaît compliqué à mettre en place. Les différentes sources de données annexes comme la BDTOP0 2015 permettent d'extraire rapidement certaines classes rapidement mais les classes d'utilisation du sol sont plus complexes à extraire.

Le temps imparti pour mener à bien cette étude et remplir tous les objectifs ne m'a pas suffi pour finaliser cette étude. Trois images du sud ouest de l'île n'ont pas encore été traitées (zone du piton de la Fournaise) et les bases de validation des images couvrant l'ensemble de l'île restent à créer pour valider le transfert des modèles de classification sur le territoire insulaire.

ÉVOLUTIONS POSSIBLES

Les sols nus apportent encore trop de confusions, à certaines dates, les seuils choisis ne permettent pas de différencier les sols nus des parcelles agricoles récoltées

et entraînent une sous-estimation des surfaces agricoles. De nombreux tests d'ajustement des seuils ont été réalisés mais aucune solution n'a été trouvée afin de séparer les sols nus des parcelles agricoles. Une étude menée en 2015 par Vaudour, Noirot-Coisson et Membrive (Vaudour et al., 2015) en métropole identifie différentes classes de sol nus et d'utilisation du sol grâce à une classification supervisée basée sur l'algorithme des SVM avec des résultats encourageants : 77% de précision globale. Il serait donc intéressant d'établir de nouveaux jeux de données incluant les sols nus et de tester cette méthode sur les images disponibles sur Saint-Pierre.

Pour les classifications postérieures, il faudra privilégier des images prises avant le début de la campagne cannière, l'idéal serait de travailler sur des images prises entre Avril et Juillet.

En ce qui concerne l'ambition de développer un modèle fonctionnant sur des outils libres, nous devons reconnaître la suprématie du logiciel eCognition en matière de segmentation. En effet, dans la littérature, la majorité des études d'analyse orientée objet d'image satellite utilise le logiciel eCognition; encore peu d'études utilisent des logiciels libres. L'appréciation de la qualité d'une segmentation est encore délicate à déterminer automatiquement et les algorithmes de segmentations développés par la communauté libre restent perfectibles. En revanche, l'ensemble des classifications des objets issus des segmentations ont été réalisées grâce à des logiciels libres.

V. CONCLUSION

L'étude présentée ici a pour but la mise en place d'une méthodologie de classification d'images satellites à très haute résolution spatiale par analyse orientée objet. Une première étude faite en 2014 propose une méthode de classification sur une zone test de Saint Pierre réalisée à partir d'une image prise en juillet 2012. Cette méthodologie se compose de trois classifications successives permettant d'extraire 8 classes d'occupation et d'utilisation du sol. La méthode proposée dans cette étude est l'adaptation à d'autres dates et à d'autres régions de l'île de l'étude de 2014. Elle permet d'extraire 9 classes d'occupation du sol grâce à :

- une première classification supervisée par règles expertes (classes réseau routier, surface urbaine, surfaces en eau, sol nu et végétation)
- une deuxième classification supervisée utilisant l'algorithme des Random Forest permet de différencier les classes générales de végétation à savoir végétation naturelle et surfaces agricoles avec une précision globale de 80%
- une troisième classification supervisée basée sur l'algorithme des SVM permet de discriminer les classes d'utilisation du sol : canne à sucre, verger, prairie et maraichage et diversification avec une précision globale atteignant 95%.

L'analyse de plusieurs images couvrant la zone de Saint-Pierre a permis de mettre en place plusieurs modèles de classification issus de l'apprentissage automatique des algorithmes utilisés : Random Forest et SVM. Ainsi, 4 modèles ont été établis à partir des bases d'apprentissage établies sur les quatre dates étudiées. Toutefois, l'application de ces modèles à d'autres images restent à valider sur l'ensemble du territoire.

BIBLIOGRAPHIE

Anderson, J. R. (1976). A land use and land cover classification system for use with remote sensor data (Vol. 964). US Government Printing Office.

Antunes, A. F. B., Lingnau, C., & Da Silva, J. C. (2003). Object oriented analysis and semantic network for high resolution image classification. *Anais XI SBSR, Belo Horizonte, Brasil*, 05-10.

Aulanier F. (2012). Méthodologie d'évaluation des services environnementaux de l'agriculture par approche spatiale à la Réunion. Paris: AgroParisTech. Mémoire d'ingénieur, spécialité 'Environnement, Eau, Déchets et Aménagements Durables', 81 p.

Baatz and Schäpe, 2000 M. Baatz, M. Schäpe (2000) Multiresolution segmentation — An optimization approach for high quality multi-scale image segmentation J. Strobl, T. Blaschke, G. Griesebner (Eds.), *Angewandte Geographische Informations-Verarbeitung XII*, Wichmann Verlag, Karlsruhe, pp. 12–23

M. Baatz, C. Hoffmann, G. Willhauck (2008) Progressing from object-based to object-oriented image analysis, Springer, Heidelberg, Berlin, New York (2008), pp. 29–42

Begué, A., Degenne, B., Baillarin, F., Lemonnier, H., Lebourgeois, V., Gargaros, D. (2005). Compte rendu de fin de recherche d'opérations d'une recherche financée par le Ministère de la Jeunesse, de l'Education et de la Recherche, le projet SUCRETTE (Suivi de la Canne à Sucre par Télédétection), CIRAD, SPOT Images, 117p

Bellon (2014). Évaluation du potentiel des images satellitaires à très haute résolution spatiale et leur analyse orientée objet pour la cartographie de l'occupation du sol à la Réunion. Mémoire de fin d'étude, Master Géosphère spécialité « Télédétection et Risques Naturels », Université de la Réunion.

Benz et al., 2004 U.C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, M. Heynen Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information *ISPRS Journal of Photogrammetry and Remote Sensing*, 58 (3–4) (2004), pp. 239–258

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1), 2-16.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. *Wadsworth & Brooks*. Cole Statistics/Probability Series, Pacific Grove, Cal.

Carleer et al., 2005 A.P. Carleer, O. Debeir, E. Wolff Assessment of very high spatial resolution satellite image segmentations, *Photogrammetric Engineering & Remote Sensing*, 71 (11) (2005), pp. 1285–1294

Darwish et al., 2003 A. Darwish, K. Leukert, W. Reinhardt Image segmentation for the purpose of object-based classification *IEEE International (3)*, pp. 2039–2041

Definiens, A. G. (2007). Definiens Developer 7 Reference Book. Definiens AG, München, 21-24.

Definiens, A. G. (2009). Definiens Developer 7 User Guide. München: Definiens AG.

Frauman and Wolff, (2005) Segmentation of very high spatial resolution satellite images in urban areas for segments-based classification. In: *Proc. International Symposium Remote Sensing and Data Fusion Over Urban Areas and 5th Intern*, 14–16 March 2005.

Grömping, U. (2009). Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*.

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6), 610-621.

Hay et al., (2005). G.J. Hay, G. Castilla, M.A. Wulder, J.R. Ruiz An automated object-based approach for the multiscale image segmentation of forest scenes *International Journal of Applied Earth Observation and Geoinformation*, pp. 339–359.

Hu et al., (2005). X. Hu, C.V. Tao, B. Prenzel Automatic segmentation of high-resolution satellite imagery by integrating texture, intensity, and color features *Photogrammetric Engineering & Remote Sensing*, pp. 1399–1406.

Johansen et al., (2007) K. Johansen, N.C. Coops, S.E. Gergel, J. Stange Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification *Remote Sensing of Environment*, pp. 29–44

Laliberte et al., (2007) A.S. Laliberte, E.L. Fredrickson, A. Rango Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands *Photogrammetric Engineering & Remote Sensing*, pp. 197–207

Lang, (2008) Object-based image analysis for remote sensing applications: Modeling reality — Dealing with complexity

Lagabriele, E., Metzger, P., Martignac, C., Durieux, L., & Grellet, G. (2005). Guide critique d'utilisation des informations produites dans le cadre du projet TEMOS à La Réunion.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *Journal of statistical software*, 25(1), 1-18.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R news*, 2(3), 18-22.

Maxwell, T. (2005). Object-oriented classification: Classification of Pan-Sharpened QuickBird imagery and a fuzzy approach to improving image segmentation efficiency. University of New Brunswick (Canada).

Zhang et al., (2005) Q.F. Zhang, G. Pavlic, W.J. Chen, R. Fraser, S. Leblanc, J. Cihlar A semiautomatic segmentation procedure for feature extraction in remotely sensed imagery *Computers & Geosciences*, pp. 289–296.

ANNEXE 1 : Détails de la nomenclature utilisée

2. Surfaces artificialisées	
2.1. Végétation naturelle	Cette classe englobe les formations végétales conservées et les espaces dégradés ou anthropisés qui ont été colonisés par des espèces végétales de façon naturelle. Y compris la végétation herbacée : Végétation de faible productivité constituée principalement de graminées et de broussailles (la plupart espèces exotiques et envahissantes). Végétation arbustive et buissonnante : formation végétale constituée principalement de buissons et d'arbustes (peux comprendre quelques arbres isolés). Y compris les friches (jachères délaissées peuplées par des espèces envahissantes caractéristiques des zones qui ont suivi une modification de l'écosystème locale). Deux espèces caractéristiques à la Réunion : Baie Rose (ou Faux Poivrier - <i>Schinusterebinthifolius</i> Raddi), Avocat marron (<i>Litsea glutinosa</i>) Forêts et formations arborées : Formations végétales denses constituées principalement d'espèces arborées. Y compris les forêts semi-sèches sur les remparts des ravines et les forêts des Hauts Végétation peu couvrante : Végétation éparse généralement présente dans des zones en dégradation ou en milieux défavorables à l'établissement d'une couverture végétale continue
2.2. Sol nu	Sol dépourvu de végétation et affleurements rocheux. Y compris les plages, les lits des rivières (fonds de ravine) et autres surfaces en sable, graviers et galets. Y compris les parcelles agricoles en sol nu : terre qui, après la récolte des cultures, reste exposée. Y compris les parcelles labourées suite à la fin de culture.
2.3. Surface urbaine	Toute surface intégrée dans un milieu urbain artificialisé. Y compris les Espaces verts urbains - Espaces végétalisés inclus dans le tissu urbain. Y compris jardins, parcs, cimetières, terrains de sport, vergers et potagers (sur des parcelles d'usage familial et non de production agricole) Y compris les Bâtiments entourés d'un tissu urbain continu et les Bâtiments éloignés des noyaux urbains, y compris les hameaux et les Serres ¹ . Sol nu artificialisé - Sol dépourvu de végétation, modelé par les activités humaines. Y compris les dépôts, les terrains hippiques, les carrières, les zones exposées suite à des travaux pour futures constructions (maisons/ résidences, serres, zones industrielles...)
2.4 Réseau Routier	Axes principaux de communication : routes nationales et secondaires de 1 à 4 voies.
2.5 Surfaces en eau	Réseau hydrographique : lits de rivière principaux et secondaires. Piscines et bassins artificiels en plein air (réservoirs, retenus collinaires...), y compris les bassins d'épuration des eaux.

1. Surfaces agricoles en culture

1.1 Canne à sucre	Parcelles agricoles cultivées en canne à sucre
1.2. Prairie	Parcelles agricoles destinées à la culture de plantes fourragères, principalement de graminées, destinées au pâturage ou à la fauche. A la Réunion on distingue deux types de prairies : les prairies tempérées et les prairies tropicales.
1.3. Maraîchage et diversification	Culture diversifiée de légumes, d'espèces aromatiques, florales et de certains fruits et céréales. Espèces principales cultivées à la Réunion : Chou-fleur, poivron, aubergine, arachide, piment, manioc, légumineuses, tomate, salade, gingembre. Y compris dans cette classe les ananas, les fraises et le maïs.
1.4. Verger	Culture d'arbres ou arbustes fruitiers (Longanier, Manguier, Avocatier, Litchi, Agrumes, Cocotier, Pitaya, Papayer, Jaquier, Grenadier). Y compris dans cette classe les Pépinières, les Vergers mixtes, les Palmeraies et les Bananiers.

ANNEXE 2 : Caractéristiques des images satellites Pléiades couvrant le territoire de La Réunion

Images date de prise de vue	Type	Bandes et résolution spatiale	Format	Zone couverte
27/07/13 commande 2014	Bundle	PAN (0.5m) B, V, R, PIR (2m)	tiff	Communes de Saint-Pierre, Saint-Louis, Le Tampon, l'Entre-Deux et Petite-île
04/03/13 commande 2015	Fusionné	B, V, R, PIR (2m)	jp2000	Communes de Saint-Pierre, Le Tampon,
28/05/13 commande 2014	Bundle	PAN (0.5m) B, V, R, PIR (2m)	tiff	Communes de Saint-Pierre, Saint-Louis, Le Tampon, l'Entre-Deux et Petite-île
04/06/13 commande 2015	Fusionné	B, V, R, PIR (2m)	jp2000	Communes de Petite-île, Saint-Joseph, Saint-Philippe, Saint-Rose, La-plaine-des- palmistes, Le Tampon, Bras-Panon
14/08/13 commande 2015	Fusionné	B, V, R, PIR (2m)	jp2000	Communes de Petite-île, Le Tampon, Saint-Joseph
20/08/13 commande 2014	Bundle	PAN (0.5m) B, V, R, PIR (2m)		Communes de La Possession, Le Port, Saint-Paul, Les Trois-bassins, Saint-Leu, Les Avirons, L'Etang-salé
27/08/13 commande 2015	Fusionné	B, V, R, PIR (2m)	jp2000	Communes de Saint-Suzanne, Saint- André, Bras-Panon, Saint-Benoit
02/09/13 commande 2015	Fusionné	B, V, R, PIR (2m)	jp2000	Communes de La Possession, Saint- Denis, Saint-Marie, Salazie, Cilaos, Le Tampon, l'Entre-Deux, Saint-Pierre, Saint- Louis, Saint-Paul, Saint-Leu
24/10/13 commande 2015	Fusionné	B, V, R, PIR (2m)	jp2000	Communes de Saint-Benoit, Sainte-Rose
04/01/14 commande 2014	Bundle	PAN (0.5m) B, V, R, PIR (2m)	tiff	Communes de Saint-Pierre, Saint-Louis, Le Tampon, l'Entre-Deux et Petite-île
11/01/14 commande 2015	Fusionné	B, V, R, PIR (2m)	jp2000	Communes de Saint-Joseph, Sainte-Rose

ANNEXE 3 : Préparation des images satellites

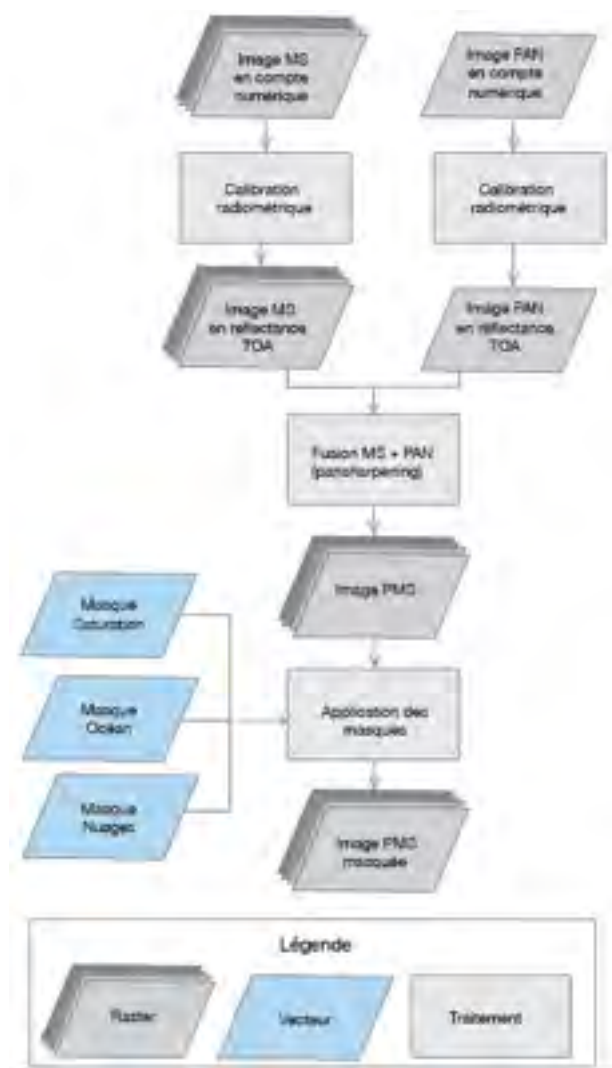


Diagramme de pré-traitement des images livrées de type Bundle

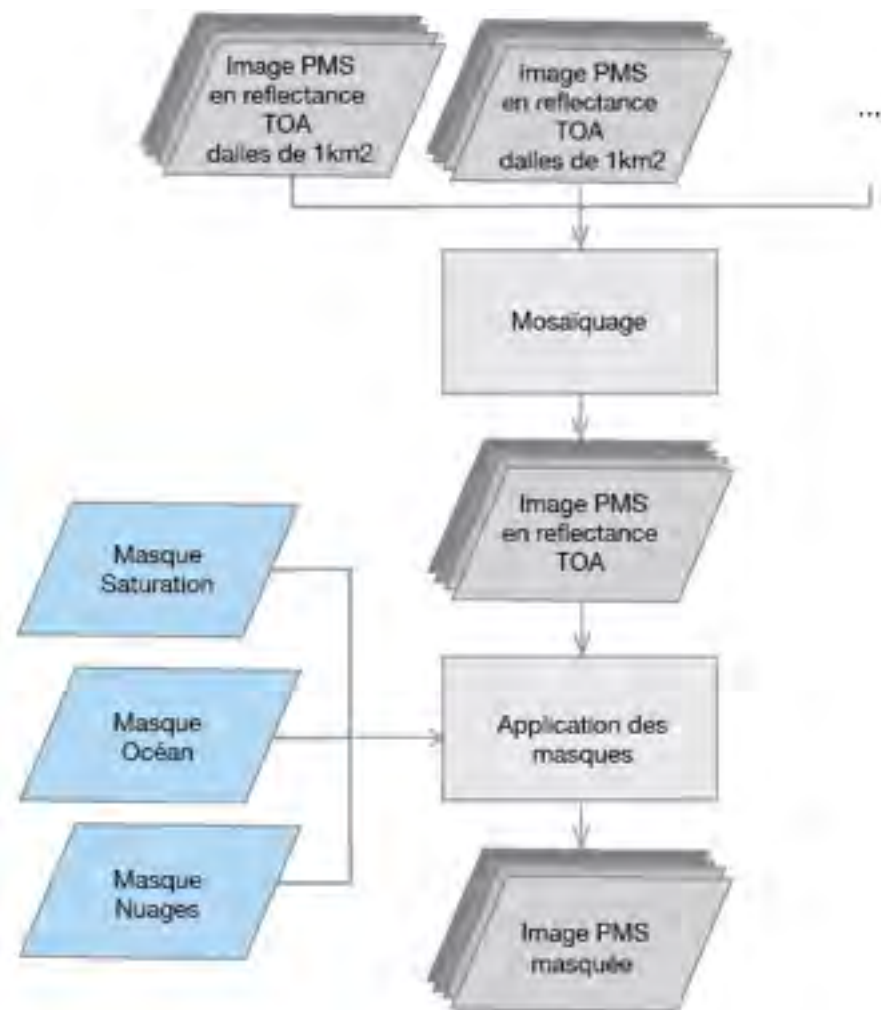


Diagramme de pré-traitement des images livrées de type fusionné

ANNEXE 4 : Informations sur les indices utilisés - Source : Bellon, 2014

T y p e	Indice	Abréviation	Formule	Description
Radiométriques	Somme de réflectances	Somme_Reflectances	$\bar{c}(v) = \frac{1}{w^B} \sum_{k=1}^K w_k^B \bar{c}_k(v)$ <p>où K est le nombre de bandes totales utilisées, k une bande spectrale, w^B la somme des poids des valeurs de brillance pour toute bande k utilisée, w_k^B le poids de brillance pour la bande k et $\bar{c}_k(v)$ l'intensité moyenne de la bande k dans l'objet v.</p>	Cet indice calcule la somme des valeurs de brillance moyenne pour les bandes spectrales (dans notre cas B + V + R + PIR).
	Indice de végétation brute	Diff_R_PIR	$PIR - R$	Indice introduit par Richardson and Wiegand (1977). Une valeur positive de cet indice souligne la présence de végétation.
	Indice de Brillance des sols	IB	$\sqrt{R^2 + PIR^2}$	Calcule la luminance des surfaces. Permet de faire la distinction entre les surfaces végétalisées et les sols nus.
	Max diff.	Max_diff	$\frac{\max_{i,j \in K_B} \bar{c}_i(v) - \bar{c}_j(v) }{\bar{c}(v)}$ <p>Avec i,j les images considérées, $\bar{c}(v)$ la valeur de brillance de l'objet v, $\bar{c}_i(v)$ (respectivement $\bar{c}_j(v)$) l'intensité moyenne de l'image i (respectivement j) pour l'objet v, c_k^{max} la valeur de brillance possible la plus élevée pour l'image k, et K_B le poids positif de la brillance dans les images.</p>	Calcule le contraste radiométrique entre les objets selon les moyennes d'intensité et de brillance des bandes spectrales (dans notre cas B + V + R + PIR).
	Indice de couleur de la végétation	ICV	$V - R$	
	Indice de couleurs froides	ICF	$\frac{R - PIR}{R + PIR} \times max\ diff$	
	Indice d'irrigation	II	$\sqrt{B^2 + V^2 + R^2 + PIR^2}$	
	Indice de la végétation verte	IVV	$((-0,290 \times B) - (0,652 \times V)) + ((0,6 \times R) - (0,491 \times PIR))$	Calcule la quantité de végétation verte

	Moyenne	Mean_X $X = \{B ; R ; V ; PIR\}$	$\frac{1}{n} \sum_{i=1}^n x_i$	Moyenne des valeurs radiométriques des pixels composant un objet dans une bande spectrale donnée (vert, rouge, bleue, PIR)
	NDVI	NDVI	$\frac{PIR - R}{PIR + R}$	Indice de végétation par différence normalisé (Tucker , 1979). Il rend compte de la biomasse des végétaux.
	NDWI	NDWI	$\frac{PIR - MIR}{PIR + MIR}$	Indice de stress hydrique par différence normalisée (Gao ,1996). Il rend compte de la teneur en eau des surfaces.
	Indice de végétation par quotient	Rapport_R_PIR	$\frac{PIR}{R}$	Birth and McVey (1968) Cet indice capture le contraste entre les réflectances dans les bandes rouge et infrarouge.
	Rapport Somme de réflectances et PIR	Rapport_Somme_PIR	$\frac{\bar{c}(v)}{PIR}$	
	Ratio (Bande X / Somme de réflectances)	Ratio_X $X = \{B ; R ; V ; PIR\}$	$\frac{\{B; PIR; R; V\}}{\bar{c}(v)}$	
	Ecart type	StdDev_X $X = \{B ; R ; V ; PIR\}$	$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$	Ecart-type des valeurs radiométriques des pixels dans une bande donnée (B,R,V ou PIR)
Texturaux	Distance euclidienne moyenne		$\frac{\sum [\sum_{\lambda} (x_{c\lambda} - x_{ij\lambda})^2]^{\frac{1}{2}}}{n - 1}$ <p>où $x_{ij\lambda}$ = la valeur radiométrique du pixel (i,j) pour une bande λ $x_{c\lambda}$ = la valeur radiométrique du pixel centrale de la fenêtre pour la bande λ n = le nombre de pixels dans la fenêtre</p>	<p>La moyenne de la distance euclidienne en une dimension, entre la valeur de chaque pixel dans une fenêtre donnée (dans notre cas 50x50) et la valeur du pixel central dans cette fenêtre pour une bande donnée (dans notre cas la bande V)</p> <p>(Irons et Petersen, 1981)</p> <p>Dans notre cas cette texture à été calculé pour la bande V.</p>
	Matrice de co-occurrence des niveaux de gris	GLCM	<p>Dans les formules ci-dessous :</p> <p>P = entrée de la matrice de co-occurrence</p> <p>$P_{i,j}$ = élément de la ligne i, colonne j de P</p> <p>N_g = Nombre de niveaux de gris de l'image</p>	<p>(Haralick, 1979)</p> <p>La GLCM permet de calculer un nombre multiple de paramètres statistiques qui décrivent différents aspects de la texture</p>

			$\mu_j = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} j P_{i,j}$ $\sigma_j = \sqrt{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{i,j} (i - \mu_i)^2}$	<p>(Clausi, 2002).</p> <p>Tous les indices ont été définis pour toutes les directions (all dir.) et calculés pour l'ensemble de bandes (B,V,R et PIR).</p> <p>Les paramètres sont présentés ci-dessous.</p>
	GLCM Contraste	GLCM_Contrast	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{i,j} (i - j)^2$	Mesure le contraste de la texture des surfaces.
	GLCM Corrélation	GLCM_Correlation	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{(i - \mu_i)(j - \mu_j)}{\sigma_i \sigma_j} P_{i,j}$	Mesure la corrélation de la texture des surfaces.
	GLCM Dissimilarité	GLCM_Dissimilarity	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{i,j} i - j $	Mesure l'hétérogénéité de la texture des surfaces.
	GLCM Entropie	GLCM_Entropy	$-\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{i,j} (\log p_{i,j})$	Mesure la dispersion de la texture des surfaces.
	GLCM Moyenne	GLCM_Mean	μ_i	Mesure la moyenne de la texture des surfaces.
	GLCM Ecart type	GLCM_StdDev	σ_i	Mesure l'écart type de la texture des surfaces.

ANNEXE 5 : Détails des Jeux de données

Images	Classe	Apprentissage 80%	Validation 20%	Total classe
28/05/2013				
1er JEU DE DONNÉES	Végétation naturelle	241	60	301
	Parcelle agricole	203	73	366
			TOTAL jeu de données	668
2ème JEU DE DONNÉES	Canne à sucre	341	92	433
	Verger	21	8	29
	Prairie	74	14	88
	Maraichage et divers	77	14	91
			TOTAL jeu de données	641
02/09/2013				
1er JEU DE DONNÉES	Végétation naturelle	192	49	241
	Parcelle agricole	277	70	347
			TOTAL jeu de données	588
2ème JEU DE DONNÉES	Canne à sucre	342	86	428
	Verger	24	7	31
	Prairie	76	20	96
	Maraichage et divers	70	18	88
			TOTAL jeu de données	641
03/04/2013				
1er JEU DE DONNÉES	Végétation naturelle	66	17	83
	Parcelle agricole	113	29	142
			TOTAL jeu de données	225
2ème JEU DE DONNÉES	Canne à sucre	238	60	298
	Verger	14	4	18
	Prairie	38	10	48
	Maraichage et divers	16	5	21
			TOTAL jeu de données	385
04/01/2014				
1er JEU DE DONNÉES	Végétation naturelle	243	61	304
	Parcelle agricole	292	74	366
			TOTAL jeu de données	670
2ème JEU DE DONNÉES	Canne à sucre	293	74	367
	Verger	24	7	31
	Prairie	36	9	45
	Maraichage et divers	56	15	71
			TOTAL jeu de données	514